DOCUMENT RESUME

ED 203 376 CS 206 464

AUTHOR Peters, Eileen
TITLE Basic Skills Improvement Policy Implementation Guide Number 2. Writing Assessment Manual. Revised Edition.
INSTITUTION Massachusetts State Dept. of Education, Boston.
PUB DATE Mar 81
NOTE 110p.

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Discourse Analysis; *Educational Assessment; Elementary Secondary Education; *Evaluation Methods; Guidelines; Holistic Evaluation; *Program Evaluation; State Programs; *Writing Evaluation; Writing Instruction; Writing Skills

ABSTRACT

Treating all major aspects of the writing assessment process, this manual is designed to help school districts evaluate writing in order to improve students' writing skills. The first major section of the manual presents background information on general measurement issues, such as evaluating writing ability by means of direct and indirect measures, the importance of validity and reliability, the value of multiple writing samples, and general comparisons of holistic and analytic scoring methods. The remainder of the manual contains detailed step-by-step guidelines for conducting a writing assessment program in accordance with the Massachusetts "Policy on Basic Skills Improvement." Guidelines and pertinent information are offered concerning creating and administering the writing assignment, using holistic scoring procedures, using analytic scoring procedures, setting minimal standards on holistically/analytically scored papers, reporting assessment results, and making instructional applications of the results. Appendixes offer additional information and forms for constructing writing tests, scoring the tests, orienting readers and aides to their tasks, recording analytically scored data, and reporting basic skills assessment results. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED203376

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

BASIC SKILLS IMPROVEMENT POLICY

IMPLEMENTATION GUIDE #2

(Revised Edition)

WRITING ASSESSMENT MANUAL

Massachusetts Department of Education

March 1981

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Massachusetts
Dept. of Education
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Copyright © 1981 Massachusetts Department of Education.

This publication is restricted for use in Massachusetts schools for the specific purpose of implementing the Massachusetts Basic Skills Improvement Policy and Regulations and may be duplicated by Massachusetts schools for that purpose. Duplication for use for other purposes requires written permission from the Massachusetts Department of Education.

CS206464

Massachusetts Board of Education

Ann H. McHugh, Chairperson
James L. Green, Vice Chairperson

John W. Bond
James P. Doherty
Millie Clements
James R. Grande
Howard A. Greis
Charles T. Grigsby
Mary Ann Hardenbergh
Armando Martinez
Edwin M. Rossman
Donald R. Walker

Gregory R. Anrig, Commissioner of Education, Secretary

Bureau of Research and Assessment

Allan S. Hartman, Director
Tracy Libros
Leslie S. May
Matthew H. Towle
Betty Hancock
Cathy Ridge
Hyla Shapiro

The Massachusetts Department of Education insures equal employment/educational opportunities/and affirmative action, regardless of race, color, creed, national origin, or sex, in compliance with Title IX, or handicap, in compliance with Section 504.

Publication #12296, approved by John Manton, Acting State Purchasing Agent

This manual was written by:

Eileen Peters
Project Director

Educational Testing Service, Northeastern Regional Office, 111 Washington Street, Brookline, Massachusetts 02146

ACKNOWLEDGEMENTS

The holistic scoring guidelines presented in this manual are based primarily on Mary E. Fowles' description of holistic scoring procedures in Education Testing Service's Manual for Scoring the Writing Sample for the Basic Skills Assessment. Criteria for judging the writing assignment are drawn from materials developed by Gertrude C. Conlon. Robert J. Jones contributed to the final chapter, "Instructional Applications." Nancy Witting edited the manuscript, and Ann Aubin oversaw its production.

The present manual is a revised, updated version of Implementation Guide #3 developed for the Bureau of Research and Assessment by the ETS Northeastern Regional Office. This revised version includes many new materials on analytic scoring and standards-setting, as well as other issues relevant to the assessment of writing skills.

The final responsibility for the contents of this manual rests solely with the Massachusetts Department of Education.

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
	Policy Requirements	1
	Purpose and Organization of Manual	2
II.	OVERVIEW OF WRITING ASSESSMENT ISSUES	3
	Different Ways to Measure Writing	3
	Advantages and Disadvantages of Different Approaches	4
	The Importance of Validity and Reliability	6
	The Decision to Score Writing Samples Holistically or Analytically	9
III.	THE WRITING ASSIGNMENT	13
	State Requirements For Topic Selection	13
	Characteristics of a Good Writing Assignment	14
	Reviewing Suggested Topics	17
	Pretesting	20
	Administering the Writing Assignment	20
	Timing	20
	Test Security and Discussion of the Topic	21
	Administering Two Assignments at the Same Sitting	21
IV.	HOLISTIC SCORING PROCEDURES	23
	For the Administrator	23
	For the Chief Reader and the Assistant Chief Reader	31
V.	ANALYTIC SCORING PROCEDURES	44
	For the Administrator	44
	For the Chief Reader	51
VI.	STANDARDS-SETTING	58
	Setting Minimal Standards on Holistically Scored Papers	58
	Setting Minimal Standards on Analytically Scored Papers	68
VII.	REPORTING ASSESSMENT RESULTS	77
	Making Holistic Scores More Comprehensible	78
	Making Analytic Scores More Meaningful	82

		Page
VIII.	INSTRUCTIONAL APPLICATIONS	85
	Activity I (Grades 4 through 12)	85
	Activity II (Grades 3 through 5)	86
	Activity III (Grades 7 through 12)	87

APPENDICES

A.	Requirements for Writing Tests (Secondary Level)	89
B.	Scoring Writing Components Holistically	90
C.	Planning the Length of the Scoring Session and Estimating the Number of Readers Needed	91
D.	For the Readers	93
E.	For the Aides	97
F.	Sample Recording Form (Analytic Scoring)	102
G.	Basic Skills Improvement Policy Annual Report Form	103

I. INTRODUCTION

POLICY REQUIREMENTS

The principal goal of the Massachusetts Policy on Basic Skills Improvement is to help students master basic skills prior to high school graduation. To accomplish this goal, the Policy requires that schools develop effective basic skills programs for elementary and secondary writing.

For the elementary grades, the evaluation of writing must occur at both the early elementary level (K-3) and the later elementary level (4-6). At least one grade at each level must be assessed, although districts may choose to include additional grades from the kindergarten-to-grade-6 range. Districts may require writing samples or objective measures (e.g., a multiple-choice test), or a combination of the two types of assessment. If writing samples are elicited, they may be scored holistically (with each piece of writing judged for total effect), or analytically (with separate components assessed for proficiency or merit). The only requirement governing scoring is that the scoring method(s), in combination with standards-setting procedures, work to identify students who have not attained the minimum achievement level established by the district. Finally, each district must report, on an annual basis, the number and percentage of students who have and have not demonstrated mastery of the district's writing objectives.

At the secondary level, the initial assessment may occur in grade 7 or 8, or the first semester of grade 9. With two exceptions, state requirements governing secondary-level writing are identical to those covering the elementary level. The exceptions are:

- A writing sample must be elicited.
- The writing assignment(s) should cover the statewide secondary level writing objectives.[1]
- If students do not meet minimal standards, they must be re-assessed at least annually until they demonstrate mastery.

---

[1] There are no statewide objectives at the elementary level; local districts develop their own.

PURPOSE AND ORGANIZATION OF MANUAL

By treating all major aspects of the writing assessment process -- topic selection, analytic and holistic scoring procedures, standards-setting, reporting, and instructional planning -- this manual is designed to help school districts evaluate writing in order to improve students' writing skills.

The first major section of this manual, "Overview of Writing Assessment Issues," presents background information covering general measurement issues such as:

- evaluating writing ability using direct and indirect measures
- the importance of validity and reliability
- the value of multiple writing samples
- general comparisons of holistic and analytic scoring methods

Anyone concerned with assessing writing achievement or interpreting results (community members as well as professional staff) should find this section useful.

The remainder of the manual contains detailed step-by-step guidelines for conducting a writing assessment program in accordance with the Massachusetts Policy on Basic Skills Improvement. In the section on the writing assignment, topics are considered in the order in which they should be addressed by school districts. For each major topic (for example, "Using Results for Instructional Planning"), applications at both the elementary and secondary levels are discussed. Or, in the event that procedures do not vary by instructional level, as is the case with standards-setting, generalized applications are specifically noted.

The manual is intended primarily for the basic skills writing coordinator -- the person(s) responsible for conducting the scoring session. We have assumed that the coordinator, or Chief Reader, will be involved in additional phases of the assessment program, such as the standards-setting process. Nevertheless, brief sections of the manual are directed to other school officials, classroom teachers, and aides. These materials are included at the end of this manual to permit easy duplication.

II. OVERVIEW OF WRITING ASSESSMENT ISSUES

This chapter presents several measurement issues that should be addressed before commitments are made to a particular method of assessing writing ability, and prior to the actual selection of instruments and scoring procedures. These issues include: the advantages and limitations of direct performance measures (writing samples) as compared with indirect measures (e.g., multiple-choice tests); the importance of reliability and validity; and general comparisons of analytic and holistic procedures for scoring writing samples. Procedural guidelines for selecting topics, scoring responses, and establishing minimal standards are presented in the chapters that follow.

DIFFERENT WAYS TO MEASURE WRITING

Writing ability can be measured directly or indirectly. A writing sample is a direct measure of a student's ability to communicate, generally under certain prescribed conditions. Direct performance measures are usually essays, although any writing sample, such as a letter or a poem, is an example of a direct measure. People generally believe that the best way to measure the ability to organize thoughts, synthesize information, express ideas fluently, and engage in other more intricate and often subtle tasks is to have students write.

Unlike direct measures of writing ability, indirect measures are designed to tap writing skills without asking the student to produce a writing sample. In measuring writing indirectly, we are asking the student to perform certain tasks that the writer does explicitly or implicitly as he or she actually engages in the craft of writing. We might, for example, give students a sentence with words missing and ask them to select the pair of words (from a series of alternative pairs) that best fits the meaning of the sentence. Or, we might present a number of sentences and ask students to identify the one that is punctuated correctly. These questions can deal with sophisticated forms as well as mechanical operations.

ADVANTAGES AND DISADVANTAGES OF DIFFERENT APPROACHES

Although there are no definite rules that tell you which assessment mode is best in particular circumstances, knowledge of the advantages and limitations of each can help you select the type of exercise, or combination of exercises, most suitable for your students and instructional circumstances. Merits and drawbacks are summarized below.

WHAT IS BEING EVALUATED?

Direct Measure	Indirect Measure
Ability to organize and synthesize information drawn from a variety of sources, and to express ideas fluently and clearly within the chosen organizational framework.	Ability to demonstrate command of conventions under controlled conditions. This measure yields rich diagnostic information. Presented with a series of items on some convention, the student is unable to avoid the assessment issue. If, on the other hand, the student is asked to write an essay, he or she can avoid words that are difficult to spell, complex sentence patterns, etc.
Ability to demonstrate control of syntactical patterns, sense of rhetorical structure, and understanding of audience.	
Ability to handle the conventions of direct discourse under authentic circumstances. (Of course, the writer can avoid conventions with which he or she is unfamiliar or about which he or she is unsure.)	Ability to demonstrate verbal facility, an ability that strongly correlates with general writing ability.

HOW ARE STUDENTS AFFECTED?

Direct Measure	Indirect Measure
Students are encouraged to organize their thoughts and express their own ideas. They are rewarded for their ability to do this. As they write, the choices they can make at a given juncture are not limited to those presented by the test developer.	Students who have trouble organizing their thoughts, or writing under testing conditions, may still be able to demonstrate command of the language and its conventions.

The actual task of having to communicate in writing (rather than demonstrating the likelihood of being able to do so) is central to student and adult life.

Students who enjoy writing and are able to organize and present their thoughts may be frustrated by the task.

HOW DIFFICULT IS IT TO DEVELOP THE MEASURE?

Direct Measure

Requires developing fewer assignments or topics. However, tasks must be clearly defined, general enough to offer some leeway, but specific enough to set limits.

Indirect Measure

Requires writing numerous questions -- a number sufficient to address all the objectives of interest. Wording must avoid ambiguities and "give-aways." For multiple-choice questions, incorrect options should embody the most likely misconceptions. Crafting technically sound objective items is a very time consuming and difficult task.

Test items selected from published materials must be reviewed thoroughly to ensure they relate to the objectives they purport to measure.

HOW DIFFICULT IS IT TO SCORE THE RESULTS?

Direct Measure

Since responses may be judged or scored differently by different readers, or by the same reader at different times, care must be taken to establish score reliability.

Procedures for training or guiding readers to score reliably are discussed thoroughly in this manual. No matter which procedure is used, rating the quality of writing samples is much more time-consuming than scoring objective measures.

Indirect Measure

Can be scored quickly. Scoring is very accurate and consistent. Assuming that the items are technically sound, that they are measuring the same general skills (i.e., editorial skills), and that the instrument is long enough, reliability should not be a problem.

WHAT ARE THE IMPLICATIONS FOR ESTABLISHING MINIMAL STANDARDS?

Direct Measure	Indirect Measure
Several standards-setting methods are discussed in this manual. The most efficient method is built into the scoring process itself. Other methods involve additional steps that may be time-consuming.	Standards-setting, using classic methods such as Nedelsky, Angoff and contrasting-groups, is relatively straightforward. Staff may be familiar with these methods, having worked with them in reading and mathematics.

HOW CAN RESULTS BE COMMUNICATED TO THE PUBLIC?

Direct Measure	Indirect Measure
Actual samples of student writing, presented along with summary score data, will make assessment results more meaningful.	Although sample test items can be presented in addition to performance data, the report may not be as interesting or informative as one that includes examples of student writing. It may be difficult to convince some of the public that objective measures really assess writing ability.

CAN RESULTS BE USED FOR INSTRUCTIONAL PURPOSES?

Direct Measure	Indirect Measure
Yes, although their value will depend on the topic and the amount of information an individual writer supplies.	Indirect measures are powerful diagnostically, assuming that the instrument has at least three to five items covering each skill (objective) of interest.

THE IMPORTANCE OF VALIDITY AND RELIABILITY

The selection of direct or indirect measures has implications for validity and reliability -- two concepts that deserve special mention.

Validity

An instrument or assessment procedure is valid if it accomplishes what is intended: i.e., if it actually measures writing ability. If our intent is to measure students' ability to write a business letter, and we administer an assignment requiring the student to write a letter applying for a

job, the measure (assignment) is valid by definition. There is an exact correspondence between our intent (what we set out to measure) and the exercise devised to accomplish the task.

Validity of Direct Measures.

Writing samples, direct measures of writing ability, will always be valid if routine care is taken in developing the topic and interpreting assessment results. You would not, for example, ask students to write a brief note to a friend if you were interested in their ability to develop a reasoned argument.

Validity of Indirect Measures.

The content validity of objective measures is established by demonstrating that test questions (or other performance tasks) adequately sample the skills we wish to measure. There are two issues here. Individual questions must address specific skills or objectives, and the test as a whole must cover the skill domain. If you decide to use an objective test to assess writing ability, all items should be carefully reviewed with reference to your writing skills objectives. The review will not be complete until you are confident that, if students do well on this test, you can be reasonably sure that they have attained your writing objectives.

Reliability

Reliability means consistency. The reliability of a measure refers to the extent to which it is consistent in measuring whatever it was designed to measure.

Reliability of Direct Measures.

If two English teachers were asked to evaluate an essay written by the same student, and were not told exactly what to look for, they might

reach very different conclusions about the quality of writing. In introducing his study of judgments of writing ability, Diederich reports:

> Teachers who have never graded a set of papers that have previously been graded by another teacher seldom realize how commonly and seriously teachers disagree in their judgments of writing ability.[1]

Based on numerous studies, Diederich's among them, we must conclude that a reliable assessment of writing samples cannot be taken for granted -- it must be achieved through systematic procedures, several of which are described in this manual. Teachers should be trained to apply a common set of assessment standards, and should practice using those standards.

Even if teachers are trained to evaluate papers reliably, score reliability is only one part of the problem. Student reliability is the other.

If a writing sample is evaluated by two teachers working independently, and both agree that the writing is less than acceptable, that particular assessment is consistent. But how much confidence can we really have in the result? Maybe the writing assignment did not appeal to the writer. Maybe the writer just had an "off" day. In other words, the student may be a better writer than this particular work sample indicates. An obvious solution is to have students produce two writing samples on separate occasions. If readers agree in their evaluations of each piece, and the appraisals of both papers are consistent, more confidence can be placed in the results.

In the real world of numerous papers and many readers, however, perfect agreement among readers is never reached. The decision to administer and score two writing assignments, therefore, has the double effect of reducing the variation in readers' judgments and minimizing the varying quality of writing from one assignment to another.

---

[1]Paul B. Diederich, Measuring Growth in English (The National Council of Teachers of English, 1974), p. 5.

Reliability of Indirect Measures.

For objective tests, reliability is usually an estimate of how close students would come to getting the same scores on a second test of the same type (for example, a parallel form of a writing skills test). If students are given the same test twice, reliability can also refer to the correspondence between the two sets of scores. Since we will ultimately be using test results to decide whether students meet minimal standards, we want to be reasonably sure that, if we administered the same test form or a parallel one, the results would be approximately the same. The more reliable the test, the more confident we can be of the results.

A number of factors contribute to error and therefore reduce the reliability of tests and test scores. A test includes only a sample of all the questions that could have been written to measure a given set of skills. Another group of questions would yield somewhat different results. For objective tests, however, scoring does not usually contribute to error. Scoring is very accurate and consistent; perfect scoring reliability can be attained.

THE DECISION TO SCORE WRITING SAMPLES HOLISTICALLY OR ANALYTICALLY

Assuming that you have decided to assess students' writing ability directly, there is another major issue that must be resolved: how should the papers be evaluated or scored?[1] Analytic and holistic scoring are two different ways of assessing the quality of student writing samples. Each has merit, but the usefulness of each varies with the nature of the writing exercise and the specificity of the evaluation information required.

---

[1]This is optional at the elementary levels; required at the secondary level.

Holistic Scoring

To score holistically is to judge a piece of writing for its total effect. Teachers of writing read papers quickly and evaluate their total impact. If the writing contains a great many errors in spelling, syntax, or organization, the reader's impression of its effectiveness will be negatively influenced. However, the discrete elements are not as important as the total expression of a student's ideas and opinions -- that is, the overall quality of the response.

At the beginning of every scoring session, the Chief Reader trains the readers (teachers of writing) in the procedures of holistic scoring. The readers discuss the topic students wrote on and then evaluate a set of training papers selected by the Chief Reader to represent the range of writing produced. They compare the papers to each other, not to preconceived ideas of a perfect paper. They assign a score of 4 (best), 3, 2, or 1 to each paper and discuss any differences in opinion. Out of this discussion of training papers arise the guidelines or rationale for scoring the responses.

Each paper is judged by at least two different readers. If the two scores are the same (4 and 4, 3 and 3, 2 and 2, or 1 and 1) or adjacent (4 and 3, 3 and 2, or 2 and 1), they are totaled to form the final score. For example, a paper that receives readers' scores of 3 and 2 has a final score of 5. If a paper receives two discrepant scores (4 and 2, 4 and 1, or 3 and 1), it must be rescored by two different readers. If the two new scores are also discrepant, the Chief Reader and Assistant Chief Reader assign a final score to that paper.

The holistic scoring procedures as outlined in print can seem complicated, but if properly used, they result in a scoring system that is extremely efficient and, as nearly as possible, reliable. If minimal standards are set prior to the training session, all work will be accomplished with ease. The major disadvantage is that holistic scores do not yield diagnostic information.

Analytic Scoring

To score analytically is to isolate one or more aspects or elements of writing and examine them individually. Every time a teacher circles a misspelled word or a grammatical error on a student's paper, he or she is scoring that paper analytically. If the elements of writing do not involve specific conventions (as spelling does), or are complex or subtle, teachers' judgments of papers will differ. To avoid this problem, it is necessary to devise scoring rubrics, or precise standards, that can be applied uniformly to each paper assessed.

Typically, a scoring rubric consists of an objective, a score scale for that objective, and mutually exclusive definitions for each point on the scale. For example:

SYNTAX[1]

3 (High)	The sentences are varied in length and structure. The author shows a confident control of sentence structure. The paper reads smoothly from sentence to sentence. There are no run-together sentences or sentence fragments.
2 (Middle)	The author shows some control of sentence structure and only occasionally writes a sentence which is awkward or puzzling. Almost no run-ons and fragments.
1 (Low)	Many problems with sentence structure. Sentences are short and simple in structure, somewhat childlike and repetitious in their patterns. There may be run-ons and fragments.

It is also important to select sample papers illustrating each point on the scale, and to have readers score and discuss those training papers before the actual scoring begins.

After consensus on rubrics and sample papers is reached, each paper is judged (generally by two different readers) on each skill or objective of interest.

---

[1]Charles R. Cooper and Lee Odell, Evaluating Writing (The National Council of Teachers of English, 1977), p. 23.

If papers are scored analytically, it is easy to identify the particular components of writing each student needs to work on. Unless the writing assignment is simple, however, analytical scoring is usually quite time-consuming. And, unless fairly precise scoring rubrics are developed, final scores will not be reliable.

III. THE WRITING ASSIGNMENT

The primary purpose of the writing assignment is to give students the best opportunity to show how well they can write, and teachers the best opportunity to evaluate that ability fairly. This chapter covers selecting and administering a direct measure of writing. Criteria for topic selection and guidelines for review and revision are presented, along with a discussion of uniform testing conditions. State requirements pertaining to topic selection and test administration are also noted.

STATE REQUIREMENTS FOR TOPIC SELECTION

There are no specific requirements governing the number, nature, or measurement characteristics of writing assignments given at the elementary levels. An adequate assessment assumes, however, that the topics have content validity, and that they meet the criteria discussed in the next section of this manual. In addition, topics should have the properties presented in Appendix A, page 89, of this manual.

Although there is no state writing test at the elementary level, a list of possible topics was sent to elementary schools, on January 5, 1981. (See "Topic Ideas for Assessing Writing at the Elementary Level.")

At the secondary level, districts may use either topics drawn from the state writing test, or topics available from other sources. The writing assignment, or combination of assignments, should elicit responses covering the full range of the secondary-level writing objectives.

Forms 1, 2, 3, and 4 of the state secondary-level writing test are currently available for local use. Forms 5 and 6 will be issued in the summer or fall of 1981. Each test form consists of two assignments -- a letter and an essay. You may use a single form intact, or choose an essay topic from one form and a letter topic from another. If you use the state test, you must administer both types of writing assignments. The choice of two similar assignments (i.e., two essays) or a single assignment (i.e., a

letter), does not represent adequate coverage of the secondary-level writing objectives. Alternatively, you may develop your own measures, or select topics from among those commercially or publicly available. You need not administer two assignments, if one elicits responses that enable you to evaluate all the writing skills in question. Any topic or combination of topics must, however, have the properties listed in Appendix A, page 89.

CHARACTERISTICS OF A GOOD WRITING ASSIGNMENT

In developing your own topics or choosing from among topics available to your school district, you should give considered attention to the following criteria:

Clarity and Specificity

The assignment should be defined, limited and clearly presented. Instructions must be clear. Students should not have to puzzle over intent; they should be told what is expected of them. All expectations should be stated explicitly. If you want students to follow business letter form, for example, this requirement should be stated in the instructions. Similarly, if you want students to draw on personal experiences to illustrate their generalizations, you should remind them of this. Instructions should always be definite in their phrasing (for example: "Pay attention to the correct form of a business letter," or, "Be sure to use complete sentences").

The topic itself should be delimited. Try to help the writer limit the subject and organize thoughts by providing an organizing framework. Use phrases such as "compare and contrast" or "briefly describe and then analyze."

Vocabulary and Conceptual Difficulty

Although you should state all your expectations, try to state the writing assignment as briefly as possible.

Restatement may sometimes be necessary to avoid misunderstanding. But, after you have restated, you might consider whether the restatement should be used without the original because the restatement does not need additional clarification.

Always try to remember two things:

- Your purpose is to assess writing skills, not reading ability; and
- You want students to spend almost all their time thinking and writing, not reading the assignment.

For these reasons, the vocabulary used and the concepts expressed in the topic should not be too difficult for the ordinary student to understand immediately.

Meaning for the writer

A good topic speaks to writers of varying ability. Good writers are challenged; they regard the assignment as interesting rather than dull and hackneyed. Weak writers are not frustrated; they are able to organize some thoughts and present them to the reader. In large measure, the quality of the writing will depend on whether the writer is presented with an interesting and satisfying task.

While you should avoid topics that will distinguish only between your best writers and the rest of the population, you should also avoid topics that are overused or those that generate trite responses. For example, "What's wrong with television?" and "What I did last weekend" are disheartenningly dull topics. They will bore writers and readers alike.

Also to be avoided are topics that are likely to elicit emotional responses. Writing assignments on politics and racial issues are, of course, classic examples. If the student takes the "wrong" political position (from the reader's point of view), the score will be either too high -- because the reader is making up for his or her own bias -- or too low because the reader has succumbed to that bias.

Other topics that adults may not consider inflammatory will evoke very emotional responses from youngsters. Some fourth-graders might be irate if they were asked to write on the subject "Why our town needs a leash law." Similarly, the subject of parentally imposed curfews may be very annoying to mature-minded junior high school students.

Type of discourse

If your objectives call for a particular type of discourse -- for example, persuasive writing -- try to devise or select topics that will truly generate the desired type of response. This is much easier said than done. Students may respond to an assignment asking them to persuade a guidance counselor to make a schedule change, for example, by offering a series of explanations about their scheduling difficulties.

Because students will often treat topics in ways you have not envisioned, you should always pretest the topic.

Bias

We tend to think of a biased topic as one that might be considered offensive. Clearly, a topic implying the inferiority of certain cultural values would be biased, as would one asking the writer to assume a stereotypic role.

But a writing exercise may be totally inoffensive and still be biased if it places any group of students at a disadvantage. For example, an assignment asking youngsters to tell about an interesting vacation they took favors the relatively affluent. A topic asking for reactions to the Super Bowl may penalize students who are not interested in sports. Other topics may not be within the experience of urban or rural youngsters, black students, children whose parents are divorced, and so on.

Be careful not to confound students' experience with their ability to manipulate the language. Have your topics reviewed by people of different backgrounds.

REVIEWING SUGGESTED TOPICS

Before final selection is made, all topics should be reviewed by several English teachers who are familiar with the students to be assessed. Reviewers should take into account the criteria just discussed, as well as the test characteristics enumerated in Appendix A, page 89. Examples of suggested topics, each accompanied by a brief critical commentary, are presented on the following pages.

TOPIC

Grades 4 - 6

Pretend that arrangements have been made for a friend to look after your pet over the weekend.

Write to your friend, explaining the details of caring for your pet. Give your friend all the important information he or she will have to know.

You will have 30 minutes to write to your friend.

Critique

This assignment is very straightforward; the instructions are clear and precise. The second sentence reminds the writer to give details, and reinforces this instruction with the phrase "all important information." The writer is told he or she has 30 minutes to complete the task.

The vocabulary is appropriate for younger writers, though you may want to substitute "plans" for "arrangements."

The topic will generate explanatory writing, with responses varying in the amount of detail provided, clarity of expression, and so forth. The topic may, however, favor children who have pets. A possible remedy is to tell children who do not have pets to pretend that they do.

TOPIC

Grades 8 - 12

If we listen to its critics, television is to blame for half the things wrong with our lives -- everything from our poor reading habits to a high crime rate. Very few people defend television, and yet it must serve some worthwhile purpose.

What values do you see in television -- to an individual, to a family, to society? Discuss one or two of these values, telling what each value is and how it benefits people. Use specific examples to support your ideas.

-Thirty Minutes-

Critique

This topic has been characterized as a cliché turned around. Yet, despite several laudable attempts to stimulate the writer, the topic may not generate the most interesting writing. That, however, is the only major criticism.

Directions limit an otherwise unwieldy subject. They tell the writer to think about a family, an individual, and society; then to narrow the task further by selecting one or two values. The question provides an organizing principle: define values and cite specific examples to support ideas.

The exercise is not biased; every writer will find a comfortable place within its conceptual framework.

After you have reviewed topics and made a tentative selection, you should write out an answer and determine whether the topic really calls for that response. You should write your response in the allotted time, just to see if it is humanly possible to do so. The topic should be revised in the light of any discoveries you make.

PRETESTING

No matter how carefully you review, there is no substitute for pretesting. You will not know if a topic evoked a range of responses until you read through a sample set of papers. Nor can you be sure that directions are clear, the topic sufficiently limited, the timing appropriate, and so on, unless you administer the exercise under normal testing conditions.

Since you will be testing all students at a given grade level, you probably will not be able to locate a comparable pretest population. You can, however, select a few classrooms at the adjacent lower-grade level, and try out your topics. Since you are assessing writing skills to determine minimal competency, the drop of one grade level would be appropriate.

ADMINISTERING THE WRITING ASSIGNMENT

Rules governing such things as timing, revision, and prewriting activities should be applied uniformly. If you want to give students the opportunity to prepare a draft, all students must have that opportunity. Similarly, if you decide to have untimed tests, all students must be given as much time as they need to complete the assignment.

TIMING

It is always hard to judge the amount of time students will need to complete a writing assignment. It is especially difficult to estimate the time needed for revision. More able writers and conscientious students may be reluctant to part with their papers regardless of the time allotted,

while those who are less able will be glad to have the experience over with as soon as possible. You will have some idea of timing if you write a response yourself; you will have a much better indication if you pretest the topic. In determining timing, remember that the objective is not to produce a publishable piece, but a writing sample that can be assessed in terms of your district's minimal writing standards.

Because the objective is to evaluate the quality of writing rather than how quickly students can write, you may conclude that students should be given as much time as they need to complete an assignment. Although untimed tests are not contrary to state guidelines per se, they pose problems for standardization. Some teachers will call a halt when most students finish and begin to grow restless, while others will not collect the papers until all students are satisfied with their work. Imposed time limitations redress such imbalances and standardize test conditions. It is probably fairer to students to time the exercise than not to do so.

TEST SECURITY AND DISCUSSION OF THE TOPIC

It is probably a good idea to keep the topic a secret until the day students are to write. While most teachers will not discuss the topic with students if you ask them not to, a few may give an assignment very much like it. The result will be, of course, a lack of uniform testing conditions. And, most unfortunately, you will not know how well students could write if left to their own devices.

ADMINISTERING TWO ASSIGNMENTS AT THE SAME SITTING

The state writing test consists of two exercises -- an essay and a letter topic. Each test is planned for 70 minutes. There is no requirement that both the essay and letter topics be administered during the same sitting, or even on the same day.

If you are giving the state test or asking students to write on two or more topics, you may want to test on two separate days. This procedure

will reduce fatigue and control for the fact that every writer occasionally has a bad day. (Be sure to provide only the topic that will be addressed at a particular testing session.)[1]

[1]A final point on the administration of writing topics. It is very important that readers not know the identity of students whose papers they may read; anonymity is essential to ensure reader reliability. This raises practical problems in assigning students the task of writing a letter. You might ask students to use a fictitious name (for example, Pat Carson, which eliminates sex bias), or if students use their own names, block out before papers are duplicated or scored.

IV. HOLISTIC SCORING PROCEDURES

This chapter describes holistic scoring procedures for both elementary and secondary levels. It is subdivided into sections for the administrator and the Chief Reader. The Chief Reader's section covers the training and scoring procedures relevant to a conventional holistic scoring session -- one in which minimal standards are set after all the papers are scored. If your district intends to set minimal standards prior to the scoring session, this chapter should be read in conjunction with Chapter VI, "Standards-Setting". It is possible to assess writing components (e.g., organization), holistically by varying the scoring procedures. This is explained in Appendix B, page 90.

FOR THE ADMINISTRATOR

Following is a summary of major administrative tasks in planning for holistic scoring.

- On the basis of the number of readers available and the number of papers to be read, determine the length of the scoring session.
- Schedule the scoring session, allowing enough lead time for preliminary preparations such as coding papers.
- Select a Chief Reader and an Assistant Chief Reader to conduct the scoring session; select aides to assist the Chief Reader in preparing for the scoring and in conducting the reading; select readers to do the scoring.
- Work with the Chief Reader to make the necessary room arrangements.
- Decide whether you want minimal standards to be set by a panel (comprised of writing teachers, administrators, parents, and others) prior to the scoring session or after the scoring is completed. (See Chapter VI of this manual.)

Why is the Administrator's Role Critical?

There are two keys to conducting a successful holistic scoring session which are the responsibility of the administrator. First, the personnel involved must be appropriate for the tasks assigned. The Chief Reader, for example, should have considerable experience with holistic scoring and should command the respect of colleagues. Readers should be experienced teachers of writing, and aides should be extremely well organized and efficient. Second, the scoring session must be planned thoroughly, allowing ample time for preliminary preparations. If personnel are carefully selected and all logistical details attended to, the scoring session will run smoothly. Participants will find the experience rewarding, and the score results will be useful for setting minimal writing standards and for planning instructional follow-up.

The administrative procedures discussed below are designed to help you plan a productive scoring session. Before you attend to these responsibilities, however, you should read the remainder of this chapter, particularly the sections covering the Chief Reader's responsibilities.

When Should Planning for the Scoring Session Begin?

Ideally, the scoring session should be planned several weeks before the writing exercises are administered. The participants, particularly the Chief Reader, must have time to prepare carefully and thoughtfully for their roles in the session. Although lead time might vary according to the size of the school system and its experience with holistic scoring, it is safe to allow two months between administering the topic and conducting the scoring session.

How Much Time and How Many Readers will be Needed?

The length of the scoring session will depend upon:

- the number of writing exercises (topics students respond to)
- the number of papers written on each topic

- the length of the average paper
- the number of available qualified readers
- the experience of the Chief Reader and the readers.

On the average, it takes a trained reader approximately 45 seconds to score a one-page piece of writing. This means that a reader should be able to score about 40 pages each half hour. Of course, you will need to allow time for training periods and coffee breaks.

Obviously, the scoring session can be shortened by increasing the number of readers. Two factors must be kept in mind, however. First, the readers should be teachers of writing who have had experience with a range of student writing ability. Second, a large reading can become unwieldy. A trained Chief Reader can usually supervise 8 to 12 readers comfortably. If the Chief Reader has conducted a number of scoring sessions, he or she might be able to manage as many as 25 readers. Larger readings probably should not be attempted unless they are supervised by experts in holistic scoring.

The exercise in Appendix C, pages 91-92, is designed to help you plan the length of the scoring session and the number of readers needed.

When Should the Scoring Session be Held?

The ideal scoring session begins in the morning, when readers are fresh, and runs through the school day. Sufficient time is allotted for discussion and breaks.

If it is necessary to schedule two or more half-day sessions, each must be planned so that all the students papers on the same topic are read on the same day. A split reading, with some papers read on one day and the remainder read on another, will reduce score reliability appreciably.

Late-afternoon sessions should be avoided if at all possible. Experience shows that, as readers tire, they score more slowly and less reliably.

Moreover, their training will not be as productive or professionally rewarding.

Who Should be Appointed Chief Reader?

The Chief Reader conducts the holistic scoring sessions and is responsible for maintaining the scoring procedures outlined in this manual. Obviously, it is advantageous to have a Chief Reader who is experienced in the holistic scoring process. But if you decide to use someone untrained in holistic scoring, choose a teacher of writing who is highly respected by his or her colleagues, who has had experience with a range of student writing ability, who is able to work firmly and fairly with others, and who fully understands the responsibilities of the Chief Reader.

Before selecting the Chief Reader, read the section of this manual that discusses the Chief Reader's responsibilities before and during the scoring session (pages 31-43).

Regardless of whom you appoint, be certain that he or she has adequate preparation time. Depending upon the complexity of the reading, three or four days will be needed to:

- read this manual thoroughly
- plan a schedule or agenda for the session
- select training papers
- train aides
- supervise the preparation of student papers for scoring
- check all arrangements carefully

Do You Need an Assistant Chief Reader?

The Assistant Chief Reader helps the Chief Reader select training papers and resolve score discrepancies; these decisions should not be made by one person. Thus, it is essential, even in holistic scoring sessions

with only seven or eight readers, that the Chief Reader have an assistant. The Assistant Chief Reader should be a teacher of writing who has the respect of his or her colleagues and who has experienced a range of student writing ability. This assistantship may be considered a training position for future Chief Readers.

Ample preparation time -- one or two days -- must be allotted for the Assistant Chief Reader to:

- plan the scoring session with the Chief Reader
- read this manual carefully
- help select training papers

How Many Aides are Needed?

Aides prepare materials before the holistic scoring sessions. At the scoring sessions, they distribute student papers and record scores. Their work is supervised by the Chief Reader.

Aides may be parents, staff members, or even students who have not been tested. They must be efficient -- both quick and thorough -- and willing to attend conscientiously to their specific instructions.

The number of aides needed depends upon the size of the reading. For eight to twelve readers at a holistic reading, you will probably need two aides -- one to distribute papers and one to check scores. For smaller readings, one aide may be sufficient.

Who Should Read and Score the Student Papers?

Readers for holistic scoring should be teachers of writing. Teachers not specifically trained to evaluate this skill tend to focus on one or two elements of writing to the exclusion of other competencies that must be assessed. As a consequence, they are less able to judge a piece of writing for its total effect, which is a fundamental requirement for holistic scoring.

Not only should the readers be experienced teachers of writing, they also should have had experience with a broad range of student writing ability. Those who teach only Advanced Placement English, for example, or teachers who provide only remedial instruction would be inappropriate choices.

Similarly, the readers' range of experience should encompass the appropriate grade levels, although some flexibility here is possible. If papers written by ninth-grade students are to be scored, for example, readers might be selected from among sixth-grade to twelfth-grade English teachers. Or, if seventh grade is the population of interest, readers could be drawn from grades four through ten.

Unlike those conducting the scoring session, readers will not need special preparation time. You will want to communicate with them beforehand, however, to explain why the scoring session is taking place and why you have chosen them to participate. Whether you hold a meeting or cover these points by memorandum, it would be helpful to give each reader a copy of "For the Readers," Appendix D, pages 93-96 of this manual.

What Kind of Room Arrangements are Necessary?

The room in which the scoring session is held should be well-lit, well-ventilated, large, and comfortable. It should have:

- a large chalkboard
- tables to seat readers comfortably -- ideally four readers to a table
- a head table for the Chief Reader and Assistant Chief Reader
- tables for the aides, who will distribute papers and check scores
- sufficient room for the aides and Chief Reader to pass quickly among tables without disturbing the readers

Diagrams of possible room arrangements are shown on page 38. You will want to discuss the possibilities with the Chief Reader before making final decisions.

When Should Minimal Standards be Set?

Minimal standards can be set either before the reading or after the scoring is finished.

If they are established before the scoring, a panel can review papers prior to the scoring session to identify papers that are minimally acceptable in content and expression. Such a panel could be comprised of writing teachers, administrators, parents, or others whom you consider qualified. These papers may then be used to establish points on the scoring scale. For example, using a four-point scale, the panel might establish the following explanation of scores:

4 - Superior
3 - Competent performance
2 - Weak
1 - Below local minimal standards

Papers will still be judged for their overall effectiveness, but their effectiveness will be evaluated in terms of a level of competence set by the panel.

Alternatively, if standards are to be set after the scoring session, papers can be compared with each other and scored on a scale of best (4) to worst (1) without regard to a minimally acceptable score. Since each paper is read twice, final scores will range from 8 (4+4) to 2 (1+1). After the scoring session, a panel can be assembled to examine the score distributions and sample papers, and set the minimal standard, or cutting score.

You will want to read Chapter VI, "Standards-Setting," before deciding when to set standards. If standards are to be set before the scoring session, this matter should be discussed thoroughly with the Chief Reader.

Summary of Administrative Tasks

Several Weeks before the Scoring Session

1. Plan the length of the scoring session, and determine the number of readers needed. Schedule the session.
2. Select the Chief Reader, Assistant Chief Reader, aides, and readers.
3. Decide whether minimal standards will be set before or after the scoring session.
4. Meet with the Chief Reader and Assistant Chief Reader to discuss holistic scoring procedures, and plan an agenda.
5. Explain to the readers the purpose of the scoring session and the role they will play.
6. Make room arrangements.
7. Make arrangements for released time, substitutes, and other personnel matters. Remember to provide sufficient preparation time for the Chief Reader and Assistant Chief Reader.

Two Weeks Before the Scoring Session

1. Meet with the Chief Reader and Assistant Chief Reader to review the agenda. Discuss the progress they have made in planning the reading.

Two Days Before the Scoring Session

1. Check room arangements.
2. Check with the Chief Reader to be sure that training papers have been selected, coded, and duplicated, and that student papers are organized and ready to be scored.
3. Double check classroom coverage for readers, availability of aides, and other details.

FOR THE CHIEF READER AND THE ASSISTANT CHIEF READER

Your scoring session will be successful if you become thoroughly familiar with the critical components of a successful reading. These are the following:

- A thorough understanding of the holistic method, including how scoring standards differ from traditional criteria for grading, the method's legitimate applications as well as its limitations, and differences between scoring standards and minimal standards. These issues are addressed throughout this manual.

- The selection of representative training papers that exemplify hypothetical score categories, and illustrate diverse approaches to the writing assignment. This chapter of the manual covers the selection of training papers in depth.

- Scoring criteria that reflect the levels of writing to be assessed, accommodate readers' diverse values, yet yield sufficient agreement to ensure the consistency of scores. Establishing scoring criteria is discussed in this section of the manual and in Appendix D, pages 93-96.

- Thorough and efficient procedures for planning and conducting readings so that participants find scoring sessions professionally rewarding rather than burdensome and unproductive. A complete description of these procedures follows. You will also want to read the previous section entitled "For the Administrator" and Appendix E, pages 97-101.

The following sections describe the Chief Reader's tasks.

Task 1: Selecting and Scoring Training Papers

What are training papers? Training papers, or range finders, are sample test papers selected to give readers an idea of the types of student responses to expect. Training papers illustrate both the four levels of writing (represented by the score scale 4 to 1) and the different approaches students can take in addressing a given topic. Before scoring begins, the Chief Reader introduces training papers to establish scoring

criteria and acquaint readers with problems they may encounter (for example, a paper that is well written but "off-topic").

Choosing the training papers is a crucial step in the holistic scoring process. It is only by considering the training papers that the readers arrive at guidelines for scoring the students' responses to a particular exercise or reach a common understanding of standards set by a panel.

How should you select papers to train readers? As soon after the test administration as possible, read enough student papers for each exercise to get a "feel" for the different ways students have responded to the assignment. Because training papers must be prepared for the scoring session, allow sufficient lead time, ideally one month, but at least two weeks.

You should skim at least 25 to 50 percent of the student responses to each exercise or writing assignment. These papers should be selected at random from different classes in your school or district.

For each exercise, select 20 to 25 papers that you think exemplify the various types of student responses you have read.

Select papers that represent, as a group, the full range of the student writing you found for each exercise. There should be some "best" (score of 4) and "worst" (score of 1) papers for each exercise, but most should exemplify average papers (scores of 3 and 2) written for this administration. When all of the test papers have been scored, there will probably be about as many 4-papers as there are 1-papers, as many 3's as there are 2's. As you would expect, there will be many more 3's and 2's than 4's and 1's. The range of writing in the 3 and 2 categories is always much wider than in the 4 and 1 categories.

For training papers avoid using the very best or the very worst paper to represent a typical 4 or 1 paper. For every score category, there is a range of appropriate responses; readers can best realize this range if they see an example of a "middle-4" paper and a "middle-1" paper. In the actual

scoring, readers will also give 4's and 1's to some papers that are slightly better and to some that are slightly poorer than the 4 and 1 training papers.

Select some problem papers. Some of the papers you choose should be ones that readers will have difficulty scoring, either because it appears that the writer did not follow the directions exactly or because the writer responded in an unusual way. For example, if the directions for the exercise say, "Give one reason for...," select at least one test paper that gives several reasons. In the training part of the scoring session, the readers will decide how to judge such a paper.

If the directions say, "Decide which proposed use of the building you support," select a paper that effectively rejects all of the proposed uses and supports none. Again, it will be up to the readers to determine how to judge such a response.

If the directions say, " Write a letter...," select papers that are written effectively but are not in letter format. Similarly, select other papers that are in perfect letter format, but that are weak in content.

Select also other papers that are unusually creative, especially if they are only marginally related to the assignment.

All 20 to 25 papers that you select for scoring in the first exercise will become the basic training papers for the holistic scoring session.

How should you prepare the training papers? Check all of the training papers to be sure they are dark enough to photocopy well. If a paper is too light, go over the writing on the copy before proceeding further. You should also delete students' names from the papers before making photocopies for the training sessions.

Have training papers photocopied. You will need one copy for each reader, as well as a copy each for both of you. Allow enough time between the selection of training papers and the scoring session to make the necessary copies of the training papers for each exercise.

Return the original copy of each training paper to the stack of student papers.

How should you score the training papers? Using the score scale 4 (best) to 1, the Chief Reader and Assistant Chief Reader should score the papers independently. When you have both scored the papers, discuss your ratings, paper by paper. If you cannot resolve a score difference, replace the troublesome paper with one on which you both agree.

How should you order the papers for presentation? Place the training papers in the order in which you will present them to the readers. First, select six papers that you both think clearly represent scores 4, 3, 3, 2, 2, and 1 for this test administration. From these six choose either a middle-3 or a middle-2 paper as your first training paper.

Put the rest of the training papers in the order in which you want to present them. Remember to keep scores mixed.

Label all of the training papers in the order in which you have decided to present them, by writing an "A," "B," "C," and so on in a large block letter in the upper right-hand corner of the paper. (If you think that teachers may interpret these letters as designating grades, begin labeling with "G.")

You will need to keep track of the scores assigned to each training paper. You can put scores on your set of training papers or work from a master list.

EXAMPLE OF MASTER LIST

Preliminary Scores
Training Papers

Paper	Score
A	2
B	1
C	3
D	2
E	4
F	3

Task 2: Preparing the Papers for Scoring

Look over all the test papers to be sure they do not identify students by name. Be careful to check letter headings, signatures, and so on. If a student's name appears, you must block it out or cover it in some way. It is very important that readers not be able to identify writers.

Task 3: Planning the Distribution System and Devising the Scoring Sheet

The Distribution System.

Determine a method of having aides mix the test papers to be scored, so that the papers from any one English class, or from any one school are not together. If papers from different schools are not mixed, the scoring will be biased. It is very important that each reader receives a random assortment of papers.

You will find it easier to control distribution if the papers are divided into groups of 15, 20, or 25, and placed into folders. At the scoring session, you may want to place the readers in two groups, one on each side of the room. That way papers scored on one side can simply be passed on to readers on the other side for a second scoring. If you decide to use this distribution system, be sure that the same readers do not consistently work as pairs. Your distribution system should suit your particular situation and should be easily understood by the aides.

The scoring sheet. In order to prevent readers from influencing each others' scoring, the readers should record scores on separate forms. Each folder of papers can contain two identical forms -- one for each reader to use in recording scores. A set of sample forms might look like the example below.

Each paper in the folder is given a letter, beginning with the letter A. Use letters to avoid confusion with scoring numbers. The reader records his or her scores directly on the form, not on the actual papers. After the first reader has scored all the papers in a folder, an aide collects the folder, removes the first scoring sheet and distributes the folder to the second reader. When the second reader has finished scoring the set of papers, an aide can easily match the two scoring sheets and circle discrepant scores.

Example: Scoring Forms

Folder #7	
Paper Letter	Score
A	
B	
C	
D	
E	
F	
G	
H	
I	
J	
K	
L	
etc.	
First Reader's Signature:	

Folder #7	
Paper Letter	Score
A	
B	
C	
D	
E	
F	
G	
H	
I	
J	
K	
L	
etc.	
Second Reader's Signature:	

Task 4: Making Room Arrangements

Specific room arrangements will depend, of course, on your facilities and the number of readers you must accommodate. Guidelines for determining the number of readers are presented in Appendix C, pages 91-92.. Once the size of the reading is known, you will want to work with your administrator in planning room arrangements.

Three sample room arrangements are presented on page 38. Whatever plan you choose, several requirements must be kept in mind:

- The room must be well-lit and ventilated.
- The room must be large enough to accommodate the necessary tables, allowing aides to circulate without disturbing readers.
- Readers' tables must not be crowded; each reader should be comfortable and have sufficient space for papers to his or her right and left.

The following equipment will be needed:

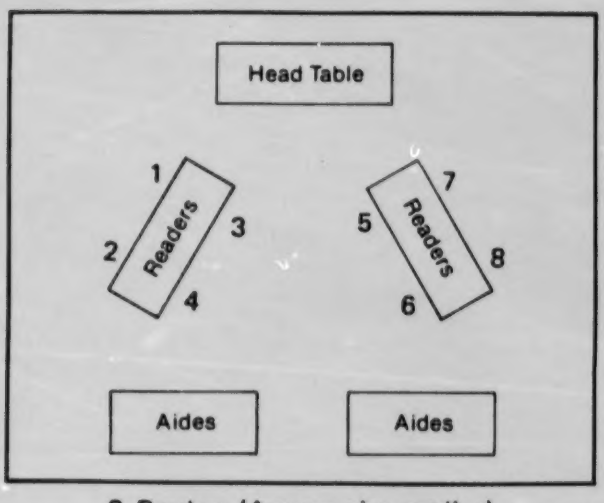
- a chalkboard
- tables to seat readers
- a head table for the Chief Reader and Assistant Chief Reader
- one table for aide(s) distributing papers and one for the aide(s) checking scores

Task 5: Planning the Agenda

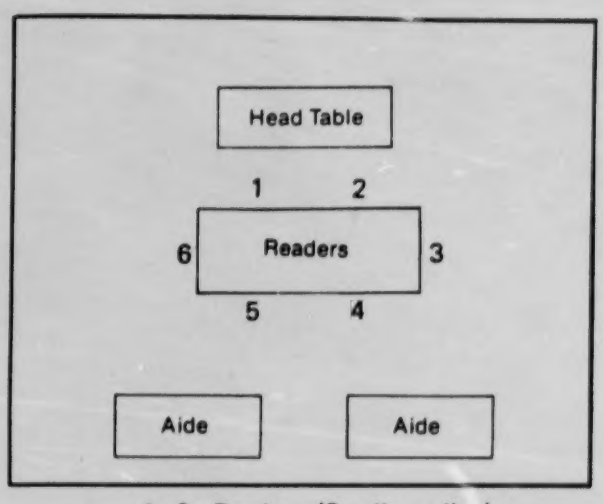
The length of a reading varies with the number of readers, the number of papers to be read, the length of the exercise the students worked on, and the experience of the Chief Reader and the readers. These factors must be considered before planning a schedule. (See the section of this chapter entitled "For the Administrator," and Appendix C, pages 91-92.

A sample schedule is presented on page 39. This agenda, for half a school day, assumes 10 readers and 300 student papers for two exercises

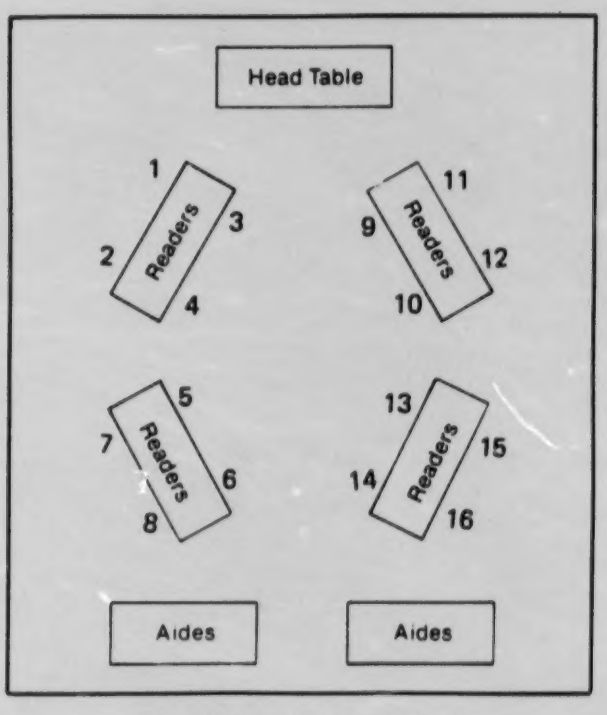
Arrangement of Tables at Scoring Session



8 Readers (Average size reading)



4-6 Readers (Small reading)



16 Readers (Large reading)

(each read twice = 1,200 readings). The sample schedule is given as an illustration, not a suggested model. Important things to note are the time allocated to training, the number of breaks scheduled, and the amount of time available for scoring as compared with the time consumed by training and breaks.

You can safely assume that it will probably take 30-40 minutes to train readers for the first exercise, with less time (approximately 20 to 25 minutes) for subsequent exercises.

Readers easily tire of sitting and reading, so you should plan a stand-and-stretch break every half hour and at least a 10-minute coffee break in the morning and in the afternoon. If the scoring of an exercise is interrupted for lunch or some other comparable length of time, you should, before resuming the scoring, present three or four new training papers to make certain the group is back on the agreed-upon track.

After you have looked over the sample agenda, you should meet with administrators to discuss: the length and number of scoring sessions, the number of readers needed, and what time of day the session(s) will be held (for example, a full school day or two afternoon sessions).

SAMPLE AGENDA
(10 Readers Scoring 300 Papers for Each of Two Exercises)

12:30 - 1:10	Introduce the first exercise and train the readers.
1:10	Stand and stretch.
1:10 - 1:40	Score papers.
1:40	Stand and stretch.
1:40 - 1:55	Score papers.
1:55 - 2:05	Coffee break.
2:05 - 2:30	Introduce the second exercise and train the readers.
2:30	Stand and stretch.
2:30 - 3:00	Score papers.
3:00	Stand and stretch.
3:00 - 3:15	Finish scoring papers.

Task 6: Taking Care of Final Details

Two days before the scoring session, do the following:

- Review the exercises and the training papers.
- Read carefully the information on conducting the reading.
- Meet with aides to be sure they understand what they are to do.
- Use the quality control checklist in Appendix E, page 101 to ensure that all materials and arrangements are in order.

Conducting the Scoring Session[1]

Training the Readers

1. Have aides distribute to readers the photocopies of the exercise (test questions and directions) about to be scored.

2. Read the directions aloud, stressing what the students were asked to do. If there are questions, keep the discussion pertinent. Mention any major problems that you have detected in the students' responses.

3. Have aides distribute photocopies of the training papers.

4. Tell readers to read Paper A quickly and consider whether the paper will probably be in the upper or lower half of the full range of papers, then consider how high (4 or 3) or how low (2 or 1) it will fall in that half. Have readers write their scores on the training paper. Of course, readers may be uneasy about trying to pinpoint a score at this early stage. You may need to reassure them that they should not expect agreement yet, and that they will re-evaluate this paper after they have read other training papers.

5. Ask the readers: "How many gave the paper a score of 4? A score of 3? Of 2? Of 1?" Then tally the scores on the chalkboard. The distribution of these early scores might look like this:

[1]If minimal standards have been set, you should read pages 60-63 of the manual before continuing further.

Readers' Scores for Training Papers

	(Score) 4	(Score) 3	(Score) 2	(Score) 1
Paper A	—	I	II	~~IIII~~
Paper B				
Paper C				
Paper D				
Paper E				
Paper F				

6. Repeat the process (read, score, tally) for the next five training papers (B to F). The scores should cluster, indicating group consensus. If the scores for any paper are spread across three or all four points of the scale, call for a discussion. Ask readers to explain why they scored the way they did. Such a paper obviously presents them with a problem that needs to be resolved. By discussing problems like these, readers should be able to establish their guidelines for this scoring session.

7. Have readers order the six training papers (A through F) from best to worst. This step reinforces the idea that each paper is to be judged in comparison with the others. Allow no talking; judgments must be independent, even in the training session.

8. Ask the readers for the order in which they ranked the papers. For a small group of readers, you might want to tally the ranking on the chalkboard to show them the extent of agreement. The tally could look something like this:

Rank Order of the First Six Training Papers

	Best	2r.d	3rd	4th	5th	Worst
Paper A				~~IIII~~	III	
Paper B						~~IIII~~ III
Paper C		I	~~IIII~~ II			
Paper D				III	~~IIII~~	
Paper E	~~IIII~~ III					
Paper F		~~IIII~~ II	I			

For this hypothetical group of eight readers, you can see that there was complete agreement about the best and worst papers. There was minor disagreement about the placement of the middle papers. This is to be expected. For every point on the scale, there is a range of papers and the range is greater for the 3 and 2 points. That is, there will probably be a variety of student papers receiving either a 4 or a 1, but a greater variety of student papers will receive either a 3 or a 2.

Because there is a range of student writing within each score category, some of the weakest 4-papers will not be much better than some of the best 3-papers. Some of the lowest 3-papers may not be much better written than the best of the 2-papers, and so on down the scale. These end-of-the-range papers may very well receive two different scores from two different readers. If the two scores are separated by only a one-point difference (4/3, 3/2, or 2/1), they are called "split scores." A split score is not only acceptable but is, for many papers, the most accurate score. When the two scores are added, the 3/2 split, for example, will result in a final score of 5. That paper has been judged better than the 2/2 papers (with a final score of 4) and not as good as the 3/3 papers (with a final score of 6).

9. Have aides distribute the next training paper (G). Repeat the process (read quickly, mark scores, tally). Continue with several more training papers, one at a time.

10. By now, readers should have reached consensus. If one or two are not scoring with the majority, you will have to identify the reasons for this and resolve the problem. The success of holistic scoring depends on the degree of inter-reader agreement. If you think the group needs more training, present a few more training papers.

11. When the readers have resolved all of the questions that you thought should be raised about scoring the test papers for this exercise, put the remaining training papers aside. You can use them if further training is necessary later in the scoring session.

12. Discuss how to handle unusually emotional or unusually creative papers, and papers that are considered off-topic. Explain that these papers are scored "0" -- meaning the reader has determined they cannot be scored. Be sure to point out that a 0 is not lower than a 1. Tell the readers to bring the 0-papers immediately to the head table for consultation.

Points to Keep in Mind

Throughout the reading you should:

- Allow no unnecessary talking among readers or aides.
- See that the papers are distributed as planned.
- Monitor the readers by reading several of the papers they have scored. If you find that a reader is not scoring according to the guidelines established in the training period, casually discuss the problem with that reader. If several readers are scoring out of line, you may need to retrain the group with a few more training papers.
- Resolve scores that are discrepant by having these re-read by two different readers. If scores are discrepant after papers are re-read, you can have them re-read a fifth and sixth time or resolve the discrepancy yourself.
- Recognize readers' fatigue and call for unscheduled breaks as necessary.

When all of the papers for an exercise have been scored two times and all discrepant scores resolved, the reading for that exercise is over.

The Next Exercise

The training process may be considerably shorter for the next exercise to be scored holistically. You will, however, follow the same procedures:

- Discuss the topic.
- Read and score the first six training papers.
- Read a few more training papers to reach scoring consensus.
- Score the actual test papers.

V. ANALYTIC SCORING PROCEDURES

FOR THE ADMINISTRATOR

The Decision to Score Analytically

To summarize previous discussions, analytic scoring refers to the evaluation of particular skills or attributes evident in a piece of writing, rather than its total effect. Whereas an holistic score represents the overall quality of a paper as compared with other papers written for the same assignment, an analytic score signifies the extent to which a student has mastered a specific writing objective or objectives. Because holistic scores are global indices, they do not yield diagnostic information in and of themselves. Therefore, at some point before instructional planning occurs, at least some holistically scored papers must be analyzed for positive features and deficiencies. The issue is not whether the writing is to be reviewed critically, but at what stage in the assessment process the analysis should occur, and how it should be accomplished.

Obviously, one alternative is to analyze all the papers at the time they are scored--to score them analytically. Before making this decision, you should review a number of factors, including:

- The need to establish score reliability
- The number and type of objectives covered by the writing assignment
- The nature of the writing assignment
- The level of diagnostic information required for instructional planning

Each of these factors is discussed below.

Need to Establish Score Reliability

In holistic scoring, the prescoring training process functions to calibrate readers -- to ensure that they will score consistently. Since

it is easier for judges to agree on a paper's overall quality (especially given only four score choices) than to concur on a paper's specific attributes, the global nature of the scoring task contributes to score reliability.

In analytic scoring, guidelines analogous to the prescoring training process must be established to ensure that readers are consistently applying the same standards as they judge a paper's particular merits. Before scoring analytically, therefore, it is necessary to develop a fairly precise set of guidelines (a scoring rubric) for each element of writing (skill objective) to be assessed. This can be a very straightforward and simple task or a complex and time-consuming one, depending upon the writing objectives and the writing assignment.

Number and Type of Objectives

Certain objectives, especially those covering conventions of standard written English, assume explicit scoring criteria. An objective focusing on correct spelling is a classic example. Any group of readers (assuming all readers were good spellers) would agree on whether or not the words in a composition were spelled correctly. Reliability would not be a problem. Spelling could be checked by one reader.

Other objectives, however, do not give such explicit guidance to readers. Take, for example, the requirement that essays be well organized. What is a well-organized paper, and how well-organized must a paper be before it is considered minimally acceptable? Because an explicit scoring standard is not embedded in the objective, it must be defined. A sample definition of a poorly organized paper follows:

> This paper starts anywhere and never gets anywhere. The main points are not clearly separated from one another, and they come in random order -- as though the student had not given any thought to what he intended to say before he started to write. The

> paper seems to start in one direction, then another, then another, until the reader is lost.[1]

Definitions of a very well-organized paper and one that was acceptably developed, with samples to illustrate each level of organization, would complete the scoring rubric for this objective.

Before making a decision to score analytically, you will want to give careful consideration to the objectives of interest, and to the task of developing scoring rubrics. You may want to consider the merits of selecting a few objectives for analytic focus, treating the remainder holistically.

Nature of the Writing Assignment

Writing assignments may be developed to assess particular writing skills, rather than to evaluate the writer's general ability to communicate. Generally speaking, the more limited the assignment (the fewer the demands it makes on students to organize and express ideas), the easier it will be to score the writing sample analytically. The reason is almost self-explanatory. Narrow or restricted assignments will cover fewer objectives requiring scoring rubrics, and the assignment will yield fewer pieces of information. For this reason, you may decide to score single sentence responses (notes or other simple messages) analytically, while scoring essays holistically. Or, you may decide to score certain writing samples (for example, letters) holistically (for content or overall writing ability) as well as analytically (for form), yielding a composite score or two separate scores. (See Appendix B, page 90.)

Instructional Planning Needs

Even though it is certainly more efficient to score essays holistically than analytically, you should not make a scoring decision without thoroughly reviewing your district's informational needs.

---

[1]Paul B. Diederich, Measuring Growth in English (The National Council of Teachers of English, 1974), p. 56.

- If you want a reliable profile of each student's writing abilities, you should score analytically.

- If you want to screen students whose overall writing ability is below your district's minimal standard, you might choose to score holistically and then analyze unacceptable papers for specific deficiencies. (You could also administer diagnostic tests of writing skills.)

- If, in addition to identifying students who are below standard and pinpointing their weaknesses, you want some general analysis of writing ability, you can score holistically and analyze sample papers at different score points.

You will notice that grade level has been omitted as a factor to consider when making a decision on scoring methodology. Except as it may influence the nature of the assignment or the objectives assessed, the writer's age should not affect the type of scoring used. For both scoring methods, the requirements and procedures for assessing the work are exactly the same for any age group.

Tasks

The analytical scoring process usually consists of several steps for establishing scoring criteria and actually scoring the writing samples. In the order in which they must be completed, these steps include:

- selecting the objectives for analytical focus

- developing a scoring rubric for each objective of interest, (defining performance at selected score points: for example, high, medium, and low)

- selecting sample papers to illustrate performance at each selected score point, for each objective assessed

- trying out the scoring criteria on a few sample papers, and possibly refining scoring rubrics on sample papers

- scoring the papers twice

- resolving score discrepancies

Some of these steps can be abbreviated or eliminated, if your focus is restricted to objectives with explicit rules of correctness. If, for example, you are scoring a business letter for correctness of form, the scoring rubric might consist of nothing more than a letter in model form, accompanied by a checklist. It would not be necessary to refine the criteria, and the scoring could be based on one review, rather than two independent readings. Most scoring criteria, however, must be refined and most papers evaluated by two readers.

Staffing Considerations

If judgment is involved in applying scoring criteria to papers, the scoring should be done by professional staff, and preferably by English teachers. If explicit rules of correctness can be provided (as, for example, a business letter format), the scoring can be done by aides. The Chief Reader should check a few papers, however, to be sure that scoring is accurate.

Time and Staff Requirements

The amount of time it will take to score the papers will depend on:

- the number of objectives to be assessed
- the complexity of the scoring criteria
- whether two independent scoring judgments are needed
- the length of the average paper
- the number of papers
- the number of readers available
- whether the scoring can be done in a single scoring session

Unlike holistic scoring, for which it is possible to provide fairly accurate time estimates, analytic scoring variables preclude such precision.

The best way to derive accurate time estimates is to score a few sample papers, using the rubrics you have developed. This should be done before deciding on the number of readers needed or the staff time involved.

You will also need a Chief Reader. Someone must develop scoring rubrics (or oversee their development), introduce the scoring task to readers, monitor the scoring, and compile the results. As in holistic scoring, the Chief Reader should be a teacher of writing who is respected by colleagues, and who is able to work firmly and fairly with others.

Before selecting a Chief Reader, you should read the section of this manual that discusses the Chief Reader's responsibilities before and during the analytic scoring session.

Depending upon the nature of the task, aides may be able to do some of the scoring, freeing the professional staff to concentrate on evaluation tasks requiring substantive judgment.

In addition, aides can help prepare materials before the scoring session and compile results after the papers are scored. If the scoring rubrics call for two independent readings of each paper, an aide's role during an analytic scoring session is identical to his or her role at an holistic reading.

Splitting the Scoring Session

Although it might be ideal to score the papers during a single session, it is not methodologically necessary to do so. (Analytic and holistic scoring differ in this respect.) Provided that the task is limited (does not include numerous objectives), it is more efficient to score in a single session. Otherwise, readers will need time to reacquaint themselves with the scoring criteria.

Advance Planning

Because there are many tasks to accomplish and several important unknowns, advance planning is critical. Do not assume that scoring rubrics can be developed quickly. Also, remember that you will not know how long it takes to score until criteria are formulated and a few sample papers are actually scored.

FOR THE CHIEF READER

Planning for the Scoring Session

Developing the Scoring Rubrics

A scoring rubric is a set of standards, explicit criteria that are used in judging some dimension of writing ability. These guidelines provide the reader with a context, or evaluation focus, by indicating what to address and what to overlook in reading a paper. A rubric covering syntax, for example, says to the reader, "Concern yourself with correct syntactical patterns, parallelism, fragments, and so on, but ignore spelling, diction, punctuation, etc."

The rubric also provides guidelines for judging the adequacy of student responses. These standards can be very simple and straightforward -- for example, the directive that a sentence be correctly punctuated. Or rubrics can be complex -- for example, defining specific attributes of a very well-developed, moderately well-developed, and poorly developed essay. The latter type of rubric should also include sample papers illustrating the categories or levels that are defined.

The difference between scoring rubrics and objectives

Although writing objectives establish the particular context for evaluation, they almost never provide explicit scoring guidelines. This is not their purpose. An objective stating that writing samples be clearly written, for example, does not tell readers specifically what to look for when scoring papers for clarity of expression. If the elements of writing (i.e., clarity or organization) do not involve specific conventions, or are complex or subtle, readers will make different judgments as they score papers. To avoid this inconsistency and improve score reliability, it is necessary to define terms as precisely as possible.

Examples of Scoring Rubrics

Following are two examples of scoring rubrics. One pertains to a modest writing assignment, the other to a more complex task.

1. A Rubric for a modest writing assignment

Suppose you have asked third-graders to write one or two sentences explaining that they have gone to a friend's house to play and indicating when they will return home. The objective of the exercise is twofold: first, youngsters must identify the pieces of information that are essential; second they must present the information in sentence form, in a way that others can clearly understand it. In order to establish scoring criteria, you must answer two questions:

- What are the essential pieces of information to be conveyed?
- As a matter of form, what properties should the sentence have?

Assume that you want the message to contain precise information on where the student has gone, why, and when he or she can be expected home. In addition, you are not concerned with spelling, word choice, and internal punctuation, but want to focus on the student's presentation of complete thoughts and use of beginning capitals and ending punctuation. The scoring rubric might appear as follows:

SCORING RUBRIC[1]

Content	Yes	No
The essential information is presented. The writer gives the name of the friend visited, indicates he or she has gone there to play, and specifies the time of return.	___	___
Form		
1. The sentences express complete thoughts, though they may be simple and perhaps somewhat awkward. Some control of sentence structure is demonstrated. There are no run-ons or fragments.	___	___
2. The sentences begin with capital letters.	___	___
3. The sentences conclude with appropriate ending punctuation.	___	___

[1]This is just an example; any number of refinements are possible. For example, you may want to establish a scale for content.

High - All essential information is provided.
Middle - Some, though not all, essential information is provided. Other information is not included, or it is somewhat vague.
Low - An insufficient amount of information is presented.

To complete the rubric, you should select sample student responses that clarify the criteria. For example, select a paper that merely lists the information and one that presents it in correct sentence form. Choose a paper that demonstrates well-constructed sentences and one that does not convey the essential information. Select examples of varying sentence quality, and so on.

2. Rubrics for complex writing assignments. Assume that you want to use analytic scoring for the following assignment, taken from one of the state writing tests for the secondary level.

> We all remember certain people from our past. You may remember someone who helped you, someone who hurt you, or someone you admire.
>
> Think about a person who made an unusually strong impression on you. Write about that person telling why you remember him or her.

Your first task is to select objectives on which to concentrate from among the state-approved writing objectives and other objectives your district may have added. As was the case for our previous example, it is not necessary to cover all the objectives in a single scoring.[1] For the sake of convenience and scoring efficiency, you may choose to cluster certain objectives into a broader category. For example, you could group spelling and punctuation together.

After objectives have been selected and categories formed, the next task is to develop a rubric for each objective or category of interest. On the following page is an example of criteria established to assess the quality and correctness of word choice.

---

[1]Of course, all the objectives must be covered eventually.

WORD CHOICE[1]

General Directions: Reward the essays for their use of good diction. Good diction consists of words that are accurate, appropriate, appealing, and direct. Consider only word choice in this score.

Possible Scores on a Five-Point Scale:

5 This score is for essays that consistently (a) use words whose form (for example, formation of noun plurals) and meaning are accurate in context, (b) maintain a level of language that balances the abstract and the concrete according to the demands of the topic, and (c) use direct and precise words to communicate clearly.

4 This score is for essays that use words (a) whose form and meaning are accurate in context and (b) whose overall style of language meets the demands of the topic. These essays do not deserve a 5-score because they are wordy (for example, "by utilization of" instead of "by using") and they need to have redundancy and "deadwood" removed.

3 This score is for essays that use words whose form is accurate, but they remain merely acceptable in their communication of meaning; that is, though the words convey an overall message, they lack precision and concreteness.

2 This score is for essays that compound the faults of a 3-score, but they deserve better than a 1-score because they use words whose form is correct. The meaning of words is vague and ambiguous and their tone inappropriate for the context. The level of communication is less than acceptable.

1 This score is for essays that (a) misuse commonly understood words, (b) invent words inappropriately, and (c) seriously hinder clear communication because they rely on vague, ambiguous language.

---

[1]Relates to the following secondary-level writing objective:

(c) Precise word choices:
- (1) Words appeal to the reader's senses
- (2) Words suit the purpose
- (3) Words are appropriate for the intended reader.

You may not want to develop scoring criteria that are as refined as the above example. There is nothing sacrosanct about a five-point scale. It is imperative, however, that definitional criteria be precise enough to permit discrimination among student responses. Readers must know exactly what they are looking for (e.g., what "word choice" means) and how to score varying levels of performance. Examples -- sample student papers -- should be used to illustrate each point on the score scale that you establish and define.

Minimal standards can be incorporated into the scoring rubric; indeed, it is most efficient to formulate scoring criteria that distinguish between minimally acceptable writing and a performance that is below standard. Therefore, before you develop rubrics, you should read pages 69 to 72 of this manual, which deal with standards-setting procedures for analytic scoring.

Finally, if you are analyzing writing according to several objectives, the mechanical process of recording scores can become quite cumbersome. Recording sheets (an example of which is included as Appendix F, page 102) will improve accuracy and efficiency.

Conducting the Scoring Session

Training the Readers

Unless readers are being asked to apply a specific convention on which teachers of English would never disagree (e.g., spelling), they must be trained to ensure that scoring criteria are clearly understood and applied consistently. It probably would not be necessary to train readers to score capitalization and ending punctuation. It is imperative that readers practice using most rubrics, such as the one presented on page 54.

Before you begin training, each reader should have:

- A scoring rubric for <u>each</u> objective on which writing will be analyzed

- Sample papers to illustrate varying levels of performance relevant to each objective, and additional samples, drawn randomly, for scoring practice
- A sample scoring sheet
- The writing assignment
- Your district's writing objectives

Start by explaining the nature of the task -- that papers are to be scored analytically to determine specific strengths and weaknesses. Then introduce the writing assignment, the objectives, and the first scoring rubric. Have readers discuss the rubric (the scoring criteria) at length. Use sample papers to illustrate the criteria and focus the discussion on the precise scoring standards that will be applied.

When you are fairly confident that readers understand the task and agree on the criteria they will apply, introduce a sample paper, and ask them to score it independently. Record scores on the chalkboard, and discuss any score discrepancies. Continue this process until score results indicate that readers are scoring consistently.

Scoring Procedures

If your scoring rubrics are complex (i.e., require considerable judgment to discriminate among score levels on a given objective), you should begin scoring at this point using the scoring rubric that readers have just been trained to use. Score all the papers on one particular objective at a time. If less judgment is required, or if the task is less complicated (e.g., determining whether a sentence expresses a complete thought), you can work through another set of criteria before having readers actually score. The point is to move as quickly as you can without burdening the readers with more criteria than they can assimilate at one time. Finally, before readers actually score, explain the recording form and the scoring procedures that will be followed.

If your scoring rubrics require little or no judgment (for example, checking punctuation), it is not necessary that papers be scored twice. In this case, simply divide papers among readers and have them rate papers and record scores. If scoring involves judgment, two readings are required to ensure reliability.

Procedures covering such things as distribution of papers in folders, student anonymity, etc., are identical to those followed in holistic scoring. The latter are presented in detail in Chapter IV.

As Chief Reader, you will need to supervise the scoring session. You should be present throughout the session to answer questions that arise and monitor the consistency of results.

VI. STANDARDS-SETTING

This chapter details procedures for setting minimal writing standards, using writing samples that have been scored holistically or analytically. Two alternative standards-setting methods applicable to holistically scored papers are presented. In the first method, minimal standards are derived before the papers are scored. The second method is implemented after all the papers are scored. Both methods involve the Chief Reader, the readers, and public participants, although the stages during which involvement occurs differ from method to method. The contrasting-groups method can be applied to holistic scores as well as to other test results. Procedures for using this method (and others that apply to objective tests) are fully explained in Implementation Guide #3 (Revised Edition): Standards-Setting, March 1981.

The issue of establishing minimal standards for analytically scored writing samples is also addressed in this chapter. Standards for each writing objective are built into scoring rubrics, and therefore developed before papers are scored. After papers are scored analytically, a single composite minimal standard is derived. The Chief Reader, the readers, and the public should be involved with both aspects of this standards-setting process.

Suggestions for public participation, using both holistic and analytic scoring methodologies, and ways of combining and describing standards conclude this chapter on standards-setting. Before attempting to set writing standards, the reader is urged to review pages 5 and 6 of Implementation Guide #3 (Revised Edition): Standards-Setting, March 1981, which cover social and political issues related to standards-setting.

SETTING MINIMAL STANDARDS ON HOLISTICALLY SCORED PAPERS

If standards are not built into the score scale prior to the reading (i.e., by designating 1 an unacceptable performance), the scores themselves will yield very little interpretive information. A paper scored 4 by two

readers (final score of 8), for example, will be among the best papers written on a given topic; it will not necessarily be a perfect paper or even an excellent one. Similarly, when compared with other papers written on the same topic, a paper scored 2 by the first reader and 1 by the second reader (final score 3) would be relatively weak. In the absence of other information, nothing about the score 3, however, suggests whether it is above or below a minimal standard.

The traditional holistic scoring process does not, in and of itself, yield minimal standards any more than a score on a multiple-choice test does. Rather, this scoring method, if used in a conventional way, yields score distributions (the number and percentage of students receiving final scores 2 through 8). It is possible, however, to establish minimal standards for holistically scored writing samples. Indeed, there are two basic methods for deriving standards. The first method is built into the scoring process itself; the second method is implemented after all the papers are scored. Each approach is discussed in detail below.

For the Chief Reader: Minimal Standards Set Prior to the Reading (Alternative Holistic Method)

Background Information

In conventional holistic scoring, the Chief Reader trains the readers to evaluate papers by comparing them with each other, not to preconceived standards of quality. Readers first decide whether a paper (when compared with other student papers) is in the upper half (the 4-3 category) or in the lower half (the 2-1 category). They then decide how well (4 or 3) or poorly (2 or 1) the paper is written. Discussion is kept to the necessary minimum, and score points are established by reading, comparing, and scoring.

You can vary this method by letting score 1 represent an unacceptable performance; i.e., writing that does not meet your district's minimal standards. Readers would still compare papers with each other, and score them as before -- 4, 3, 2, 1. But score 1 would be reserved for papers that were not only relatively weak (compared with other papers), but also unacceptably written (below your district's standards).

In order to score reliably, readers must be trained to discriminate between unacceptable and minimally acceptable writing, as well as to recognize other performance levels. Your most difficult task will be to train readers to make fine distinctions, using the lower end of the score scale. And to accomplish this, you will want to vary your training session somewhat.

Standards-Setting Procedures

Procedures are presented below in the order they should be implemented. While reading them, you may want to refer back to Chapter IV, "Holistic Scoring Procedures."

Before the Training Session

Pull more "lower-half" than "upper-half" papers. Include at least 20 lower-half papers in your sample set. Select a few papers that are very poorly written, but more that show both major weaknesses and some redeeming qualities. Choose a number of samples that you and your Assistant Chief Readers regard as borderline (2/1).

Meet with your readers to review the lower-half papers you have pulled. Consider these sample papers with reference to your district's writing objectives. Ask the group to identify papers considered unacceptable and minimally acceptable.

Keep a record of how papers are classified, along with the reasons given for each judgment.

Conduct a review session with a committee of public representatives. Present your district's writing objectives, the topic on which students wrote, and the papers you and the readers considered unacceptable and minimally acceptable. Be prepared to explain why you classified papers as you did. Also be prepared to introduce additional lower-half papers, in the event that the panel takes issue with your proposed standards. Continue the discussion and review process until a consensus is reached on examples of unacceptable and minimally acceptable writing performances. The papers the panel agrees are unacceptable become the training papers for score 1. Keep a record of how papers are classified, along with the reasons given for each judgment. The materials you will need for this session are listed on the following page.

HOLISTIC SCORING
(Alternative Method)

Materials Needed by Each Member of the Review Panel, Considering Proposed Criteria for an Unacceptable Score

- A set of your district's writing objectives
- The writing assignment
- A few sample papers that the readers think illustrate minimally acceptable and unacceptable writing. Student names must be removed.
- Additional lower-half papers to use as necessary
- A rationale for the proposed standard. The rationale might take the form of a written analysis or critique of a few papers at, above, and below the recommended standard.

During the training session. Before you begin training, tell the readers that they will use score 1 to represent an unacceptable writing performance. Explain the community review procedure that has taken place. Introduce the objectives and a range of four or five lower-half papers. Discuss these papers with the readers, focusing on papers that the review panel classified as minimally acceptable and unacceptable. Continue introducing papers until the readers' evaluation of writing performance corresponds to the review panel's classification of papers.

Now proceed as if you were running a conventional training session. Introduce a few papers representing the range of scores readers will use. Ask the readers to read all the papers, sort them (upper-half and lower-half), and score each paper. Continue presenting papers until readers are scoring consistently, and agreeing as to which papers should be scored 1.

Before scoring begins, ask the readers if they are comfortable with the unacceptable category, and if they are ready to score. If they seem unduly hesitant, continue to work with lower-half papers. Finally, tell the readers to assemble papers the group has decided to score 1. Ask them to re-read these papers, and remind them that score 1 is to be reserved for

similar, unacceptable papers they encounter. Also mention that some readers may never see 1 papers. Now begin the scoring.

After the reading. When the scoring is complete, you will have three categories of papers:

- Above minimal standard (final scores 8 - 4)
- Below minimal standard (final score 2)
- Borderline (final score 3)

Your final task is to make a desision about the borderline category -- papers that one reader judged acceptable and the other reader judged below your district's minimal standard. Several "decision rules" are listed below; others may occur to you. The only requirement is that the decision rule be reasonable, and that it be applied uniformly to every paper in the borderline category.

POSSIBLE DECISION RULES FOR BORDERLINE PAPERS

- Treat the borderline score 3 as a discrepancy, and have the papers re-read until disagreement as to mastery status is resolved.
- Treat the borderline score 3 as below (or above) the cutting score.
- Re-test students who obtain borderline scores.
- If another writing exercise has been administered, and the student's score on that assignment was above standard, classify the student as above standard. If the score on the other exercise was below standard, classify the student as below standard.

For the Chief Reader: Minimal Standards Set After Scoring (Conventional Holistic Method)

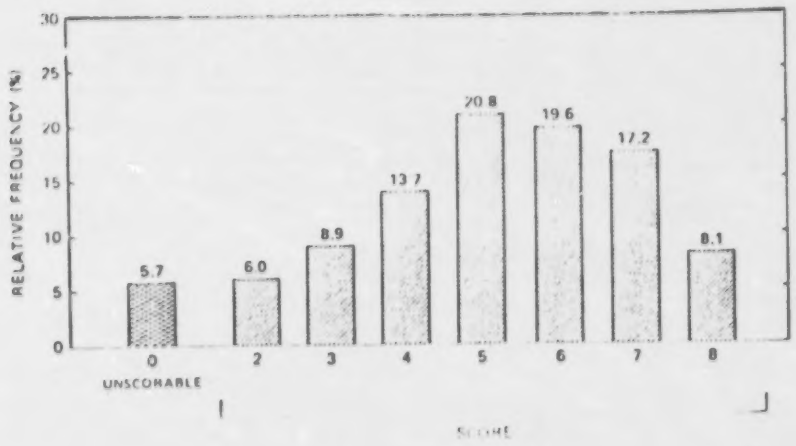
Background Information

If you choose not to incorporate standards-setting into the holistic scoring process, you will have to establish standards after all the papers are scored.

The approach to standards-setting, described below, involves a systematic examination of student papers using the writing objectives as criteria. In addition to a minimal standard or cutting score, the method will yield specific information about students' ability to deal with the various components of writing -- syntax, correctness of information, and so on. If you choose this approach, it should be implemented as soon after the scoring session as possible. The Chief Reader and the readers should be involved in the process; the public can also be involved at this stage.

Standards-Setting Procedures

Determine final scores. After each paper has been read twice, add the two readers' scores to get a final score for each paper. The score will range from 2 (1+1) to 8 (4+4). Then tally the number and percentage of papers in each score category (2 through 8). The resulting score distribution might look something like the following histogram:



Sort Papers. Place the papers in seven piles according to their final scores. (Eliminate unscorable or 0 papers; these students must be reassessed.)

Review with Readers. Tell the readers that you are going to set minimal writing standards by examining some scored papers analytically. Before you examine papers, however, have the group review the writing objectives, the writing assignment, and students' grade level(s). Discuss what your district means by a minimally competent writing performance. (You may wish to tape-record this session for future reference.)

Re-examine the papers analytically. Start with the lowest score category (papers with final scores of 2); introduce a sample middle-2 paper; ask the readers to comment on the writing relative to the writing objectives in question. After the discussion is over, ask the readers, "How many think this paper represents a minimally acceptable writing performance? How many do not?"

Read several more 2 papers, being sure to include middle-2 and high-2 papers. In each instance, discuss the paper analytically in relation to your district's objectives before calling for a summary judgment regarding minimal acceptability.

After the range of 2 papers has been covered (approximately 10 to 15 papers read), consider papers in the next highest category -- final score 3 (2 given by one reader, 1 given by the other reader). Proceed exactly as before discussing approximately 10 to 15 papers thoroughly and deciding whether each paper meets minimal standards.

Set the proposed cutting score. The process of thoroughly examining groups of papers and judging their acceptability should continue (up the score scale) until a clear majority of readers (perhaps 65 or 75 percent) rate almost all the papers in a given score category acceptable. The table on the page 66 will clarify the procedure and suggest how the cutting score should be interpreted.

ANALYTICAL RATING OF WRITING SAMPLES BY TEN READERS

Paper	Final Holistic Score	Number of Readers Rating Paper Minimally Acceptable
101	2	1
014	2	0
091	2	3
217	2	2
240	2	1
017	2	1
026	2	0
011	2	2
317	2	0
283	2	1
118	3	2
215	3	3
031	3	2
038	3	1
185	3	4
189	3	3
210	3	4
028	3	2
177	3	3
022	3	2
177	4	7
121	4	8
044	4	8
049	4	7
001	4	9
246	4	7
150	4	6
161	4	8
019	4	9
209	4	9

In this example, almost all the readers thought that papers receiving final scores of 2 or 3 did not meet minimal competency standards, while a large majority thought that the last ten papers, each with a final score of 4, were minimally acceptable.

The cutting score would thus be set at 4. Students with final scores of 4 through 8 would meet the minimal writing standard for the writing exercise in question. Students with final scores of 3 or 2 would not meet minimal standards on this exercise.

Using the sample score distribution presented on page 64, a cutting score of 4 would place 14.9 percent below the minimal standard.

Review the proposed standard. As noted above, community representatives may be included on the panel that sets the standard. If you choose not to provide for public participation at that stage of the standards-setting process, the Chief Reader's final task is to prepare for a community review. A separate set of materials will be needed for each writing exercise or topic.

HOLISTIC SCORING
(Conventional Method)

Material Needed by Each Member of the Review Panel

- A set of your district's writing objectives
- The writing assignment
- A summary of scores (the number and percentage of papers with final scores of 2, 3, 4, 5, 6, 7, or 8)
- A few sample papers that represent each level of writing on the score scale (papers with final scores of 2, 3, 4, 5, 6, 7, or 8). Be sure students' names are not on the papers.
- The proposed standard or cutting score recommended by the readers
- A rationale for the proposed standard. The rationale might take the form of a written analysis or critique of a few papers at, above, and below the recommended standard. If, for example, the proposed standard is 4, the Chief Reader should analyze a few 3, 4 and 5 papers, point out the general strengths and weaknesses of each set of papers, and explain why 4 is the proposed cutting score.

For the Chief Reader: Setting Minimal Standards on Analytically Scored Papers

Scoring rubrics, applied in the analytic assessment of writing ability, can be used to explicate minimal standards. The rubric for word choice (presented in Chapter V) posits the following criteria for scores 1 - 5.

1 This score is for essays that: (a) misuse commonly understood words, (b) invent words inappropriately, and (c), seriously hinder clear communication because they rely on vague, ambiguous language.

2 This score is for essays that compound the faults of a 3 score but deserve better than a 1 score, because they use words whose form is correct even though their meaning is vague and ambiguous and their tone inappropriate for the context.

3 This score is for essays that use words whose form is accurate, but they remain merely acceptable in their communication of meaning; that is, though the words convey an overall message, they lack precision and correctness.

4 This score is for essays that use words (a) whose form and meaning are accurate in context and (b) whose overall style of language meets the demands of the topic. These essays do not deserve a 5-score because they are wordy (for example, "by utilization of" instead of "by using") and they need to have redundancy and "deadwood" removed.

5 This score is for essays that consistently (a) use words whose form (for example, formation of noun plurals) and meaning are accurate in context, (b) maintain a level of language that balances the abstract and the concrete according to the demands of the topic, and (c) use direct and precise words to communicate clearly.

On this rubric, score 3 connotes an acceptable, though in some ways deficient performance, whereas scores 2 and 1 represent a seriously flawed writing performance.

Incorporating minimal standards in the rubric

The most efficient procedure is to build a minimal standard into each scoring rubric. Assume, for example, that your district decided that the definitions for scores 1 and 2 for the word choice rubric represented unacceptable performance levels. A minimal standard can be established by stating that, for scores 1 and 2, the level of communication does not meet the minimal standard.[1] Thus:

[1]Remember that two readers will assign scores so that final scores of 2 (1 + 1), 3 (2 + 1), and 4 (2 + 2) will be below standard. This rubric and the hypothetical standard are presented for illustrative purposes only. We are not suggesting that a 5-point scale be used, or that the bottom two scores be designated as below standard. We are recommending that some point on a score scale be designated as below standard, and that this and other points on the scale be defined in accordance with your district's expectations for writing performance.

1 This score is for essays that (a) misuse commonly understood words, (b) invent words inappropriately, and (c) seriously hinder clear communication because they rely on vague, ambiguous, non-standard language. The level of communication does not meet the minimal standard.

2 This score is for essays that compound the faults of a 3 score but deserve better than a 1 score because they use words whose form is correct even though their meaning is vague and ambiguous and their tone inappropriate for the context. The level of communication does not meet the minimal standard.

Establishing a single composite standard

In our previous example, all students whose essays received scores of 2 or 1 would be below standard on this particular writing skill. But students would also have ratings on other objectives, and these scores might or might not be below standard. Therefore, after you develop rubrics (that include standards for each objective considered independently), you will need to develop a composite standard -- one that takes performance on all the assessed objectives into account. This final standard is an answer to the question:

> How many acceptable ratings (on individual objectives) must a student receive in order to be considered minimally competent?

There is, of course, no correct answer to this question. One district may decide that students should demonstrate minimal mastery of all the district's objectives. Another may conclude that the attainment of certain objectives are essential, but that performance in one or two other skill areas could be deficient and the student still considered a minimally competent writer. Although answers, and thus standards, may vary from district to district, the standards-setting process should not be an arbitrary one. You should have sound pedagogical reasons for deciding that performance in certain skill areas is more important than in others. A standard calling simply for the attainment of 70 percent of your objectives is not defensible in and of itself. (See pages 1 and 2 of Implementation Guide #3 (Revised Edition): Standards-Setting, March 1981.

Reviewing the proposed standards

Both the scoring rubrics -- the one incorporating minimal standards for specific objectives and the proposed final composite standard -- should be reviewed by community representatives. To be safe, this review should take place before papers are scored. Then, if reviewers take issue with your rubrics (e.g., think that you should identify additional performance levels), you will be able to make the adjustment before scoring.

Theoretically, once rubrics are accepted, the review of the proposed composite standard could take place before or after the papers are scored. (Student scores on particular objectives will not be affected by the number of attained objectives your district sets as the minimal requirement.) Unless, however, you envision two district review panels -- one to review rubrics and one to consider the proposed composite standard -- it is probably more efficient to conduct a single review session. This should take place after rubrics are developed, but before papers are scored. Each member of the review panel will need a set of the materials enumerated below.

ANALYTIC SCORING

Materials Needed by Each Member of the Review Panel Considering Scoring Rubrics and Composite Minimal Writing Standard

- A set of your district's writing objectives
- A scoring rubric for each objective for which writing will be assessed analytically
- Sample papers illustrating levels of writing on each score scale to be reviewed (the same papers might be used to illustrate scale points for different rubrics). Remember to remove students' names.
- The writing assignment
- The proposed composite standard recommended by the readers
- A rationale for the proposed standard

The chart on the following page shows various ways the public can be involved in standards-setting processes. These are recommendations, not rules or requirements.

SUGGESTIONS FOR PUBLIC PARTICIPATION

	Conventional Holistic (standards set after papers are scored)	Alternative Holistic (standards set before papers are scored)	Analytic (standards incorporated into rubrics and composite standards set)
Selecting sample papers at different score levels	Observer	Observer	Observer
Developing rubrics	Observer	Observer	Observer
Reviewing rubrics	Participant	Participant	Participant
Training readers	Observer	Observer	Observer
Scoring papers	Observer	Observer	Participant
Setting the proposed cutting score	Observer	Observer	Observer
Reviewing the cutting score	Participant	Participant	Participant

Describing the Writing Standard

Assessment results (the number and percentage of students above and below standard) will not be meaningful unless the standard itself is described.

For multiple-choice tests, the standard is described simply as the number of test questions that minimally proficient students had to answer correctly. Similarly, if a single writing topic were administered and responses scored analytically, the standard would be the number of objectives which students were required to master. If these responses were scored holistically, the minimal standard is described as the final score level (2, 3, 4, etc.) that minimally competent writers had to attain.

While this is very straightforward, the introduction of more than one assessment task (i.e., two writing assignments or one writing assignment and an objective test) poses a complication. After a cutting score or standard for each task is established, a final standard must be derived and described. Four possible approaches to this problem are explained below. Each approach is followed by illustrative examples, describing hypothetical standards. In each instance, the first example pertains to the elementary level (assuming two assessment tasks have been administered). The second example relates to the use of the state secondary-level writing test.

A combined score. The two scores can be added together to form a combined score, and students can be required to reach that combined score to be considered above standard. There are two problems with this approach. First, students might do very well on one topic and very poorly on the other. Second, if a student does not meet minimal standards and has to be retested at a later date, he or she would have to be tested on both writing topics and both would have to be scored, creating more administrative work for the district. Also, by looking at a combined score rather than individual scores for the different topics, the district loses information that could be of instructional value.

DESCRIPTIONS OF HYPOTHETICAL STANDARDS	
Elementary	Students are required to receive a minimum combined score of 6 out of 16 (on two topics requiring paragraph responses) to meet local district standards.
Secondary	Students are required to receive a minimum combined score of 7 out of 16 on the letter and essay to meet local district standards.

An average score. The two scores can be averaged to form one required cut-off score and students can be required to reach that average score to be considered above standard. This approach has the same limitations as the combined score.

DESCRIPTIONS OF HYPOTHETICAL STANDARDS	
Elementary	Students are required to receive a minimum average score of 12 out of 30 (possible score points on two sentences scored analytically) to meet local district standards.
Secondary	Students are required to receive a minimum average score of 3.5 out of 8 (on the letter and essay) to meet local district standards.

A single score on one task or the other. The two scores can be kept separate, and students can be required to reach the minimum score on one writing task or the other. The main problem with this approach arises if the tasks relate to different writing skills or objectives: the student could perform well on some objectives, very poorly on others, and still be considered above standard. The student's achievement on some of the objectives might not be fully assessed.

DESCRIPTIONS OF HYPOTHETICAL STANDARDS	
Elementary	Students are required to receive a minimum score of 3 out of 8 (on an essay) OR 25 items out of 40 (on a locally developed writing skills test) to meet local district standards.
Secondary	Students are required to receive a minimum score of 4 out of 8 (on the letter) OR 3 out of 8 (on the essay) to meet local district standards.

A double score requirement. The two scores can be kept separate, and students required to reach a minimum score on each of two writing tasks. This method has three advantages. First, it ensures a thorough coverage of the writing objectives at the secondary level. Second, if a student performs above standard on one task and below standard on the other, follow-up instruction can be more directed -- the district need focus on only a limited range of objectives. Third, there is a major administrative advantage to this approach. If a student reaches the cut-off score on one task, he or she will have to be re-assessed only on the other task, rather than on both tasks. This will save testing and scoring time. Because this method is the most defensible in terms of curriculum and represents the most efficient use of district resources, it is recommended for the secondary level by the Department of Education. The double requirement is equally appropriate for elementary levels if two assessment strategies are used, and each relates to different sets of objectives.

DESCRIPTIONS OF HYPOTHETICAL STANDARDS	
Elementary	Students are required to receive a minimum score of 3 out of 8 (on an essay) AND 3 out of 8 (on another essay) to meet local district standards.
Secondary	Students are required to receive a minimum score of 4 out of 8 on the letter AND 3 out of 8 on the essay to meet local district standards.

A cautionary note on combining standards. The following types of scores cannot be added or averaged:

- holistic scores and objective test scores
- holistic scores and analytic scores
- analytic scores and objective test scores

If you have these score combinations, you must require that students reach both standards, or one standard but not the other.

VII. REPORTING ASSESSMENT RESULTS

Two major issues are treated in this chapter: state reporting requirements are summarized, and ideas for reporting assessment results meaningfully to parents and the public are included. The latter topic covers the presentation of both holistic and analytic assessment results in ways that generate interest and permit valid interpretation.

STATE REPORTING REQUIREMENTS

By August 31, 1981, and no later than August 31 of each subsequent year, school districts are required to report the number and percentage of students who have achieved, have not achieved, or are exempt from achieving minimal standards in writing. For example, if you set your standard or cutting score at 4 (on an 8-point scale) on one essay scored holistically, you would report the number and percentage of students who score 4 to 8 as above standard. Students scoring 3 or 2 would be reported as below standard.

This information must be reported for each grade level included in your district's Basic Skills Improvement Plan, and aggregated by student sex and race/ethnicity. If students are tested twice during a given school year, the final assessment results are the ones used to determine the number and percentage of students above and below standard. A sample reporting form is included in Appendix G, page 103 of this manual.

In addition to reporting results directly to the state, districts must provide assurances that the Annual Report will be released and made generally available to the public, and that reasonable steps will be taken to ensure that the public is informed of the content and availability of the Annual Report.

Finally, it is certainly important for parents to know whether their children scored above or below standard, and to be aware of the minimum

score requirement in your school district. You should also consider presenting a more comprehensive report -- one that will place the writing assessment in context and facilitate score interpretation.

MAKING HOLISTIC SCORES MORE COMPREHENSIBLE

Holistic scores are certainly more meaningful if they are reported along with representative samples of student writing, and they are most meaningful if you provide writing samples for different scores along with written commentaries.

A possible format might include a copy of the writing assignment and the objectives, followed by a brief discussion of how the papers were scored. A score distribution, showing the number and percentage of students obtaining each final score, would be reported next, as illustrated below.

SCORE DISTRIBUTIONS BY GRADE LEVEL ON EXPOSITORY WRITING
(All percentages rounded to nearest whole percentage)

Score	Grades					
	3-4 (Snowman Essay)		5-6 (Pet Letter)		7-8 (Moon Report)	
	(N)	%	(N)	%	(N)	%
8 (4/4)	(25)	10	(31)	13	(17)	9
7 (4/3)	(25)	10	(31)	13	(21)	11
6 (3/3)	(52)	21	(41)	17	(30)	16
5 (3/2)	(32)	13	(59)	25	(42)	22
4 (2/2)	(77)	31	(35)	15	(37)	20
3 (2/1)	(14)	5	(25)	10	(35)	18
2 (1/1)	(24)	10	(18)	7	(8)	4

An analysis of writing samples would follow. Explanatory and cautionary remarks should precede this discussion, for example:

The English department selected several papers representing a typical performance at several score points. The Chief Reader made final paper selections and commented on each one with reference to the district's writing skills objectives. In the pages that follow, each student's paper is presented exactly as it was written. No revisions, deletions or corrections have been made. It is important to note that none of the essays is considered a perfect or model piece of writing. Rather, each writing sample is illustrative of a range of writing ability. Among the papers scored 2, for example, some were better than the 2-paper presented in this report, while several were not as well written. As a group, the papers represent the quality of writing students actually produced, in a very limited time, without preparation, and without the chance to revise their initial drafts on the basis of teachers' comments.

If you do not have time to select and critique a sample paper representing each final score level, choose papers at score points 8, 6, 4 and 2. Or, if this is still unwieldy, concentrate on a paper that is minimally acceptable as well as one that does not meet your district's standards. For your reader's convenience, print each writing sample and relevant commentary on facing pages. An example follows.

Final Score 2
(1/1)

"I think that T.V. is good for you because if you don't have anything to do you can watch T.V."

"I think that T.V. is good for you in different ways because you mite get sick of doing nothing."

"Lots of people say T.V. is bad but their realy good chanels on T.V.

REVIEWER'S COMMENTS

Score 2

The essential weakness of this paper is the apparent inability of the writer to move in any direction away from the given question, "What good things do you see in television?" Perhaps the writer was overly impressed with the underlined "good" and "you" in the question, for of the three sentences that comprise the essay, two of them focus on the "you" ("I") and the "good," and the third reiterates the "good." Beyond these simple general statements, the only development is an illogical statement of cause: Television is good because you can watch it if you have nothing to do.

Aside from lack of substance, the writer exhibits only minimal control over mechanics, i.e., spelling: "realy," "mite," and the curious framing of the sentences in quotation marks. On the other hand, the essay does suggest that the writer has some control over conventional syntax since the sentence patterns (complex and compound: S - V - O) are complete and comprehensible.

When ranked with other papers in the sample, this essay was judged to be of low quality and was scored a 2.

MAKING ANALYTIC SCORES MORE MEANINGFUL

Whereas the inclusion of sample papers scored holistically will convey the true meaning of assessment results, papers illustrating analytic score levels will serve to confuse rather than to clarify. (A full range of examples covering ten objectives and five score levels would require fifty sample papers!)

Rather than presenting representative papers for comparative purposes, you should consider reporting scoring rubrics along with individual student profiles. The profile would consist of a total score and a score on each dimension of writing assessed in your district. The rubrics (definitions of each score point for each objective) would enable parents to interpret performance qualitatively. For reporting purposes, you might want to summarize or abbreviate the definitions you used in scoring. If you do not think it is necessary to give all this information, reduce your profile to two score categories, above and below standard, for each dimension or objective. Then, instead of presenting complete scoring rubrics, simply define an unacceptable performance level for each objective.

Two examples of completed forms follow. The first would be attached to the set of rubrics used in scoring; the second would not provide rubrics, but would call for attached definitions of below-standard performance for each skill dimension. Of course, the content of either form will depend on the objectives or components of writing that you assessed.

WRITING SKILLS ASSESSMENT

Score Profile

STUDENT NAME Alice Taylor

GRADE 8

	5	4	3	2	1
Organization	☐	☒	☐	☐	☐
Clarity and Consistency of Purpose	☐	☒	☐	☐	☐
Word Choice	☐	☐	☒	☐	☐
Sentence Structure	☐	☐	☒	☐	☐
Grammar	☐	☐	☐	☒	☐

The definitions of scores (5-1) for each objective are attached to this report. Your child received a total score of 16 out of the possible score of 25. This score placed your child (above) below minimal standards on this writing task.

WRITING SKILLS ASSESSMENT

Score Profile

STUDENT NAME Mark Thomas

GRADE 8

	Performance Level	
	At or Above Standard	Below Standard
Organization	✓	
Clarity and Consistency of Purpose	✓	
Word Choice	✓	
Sentence Structure	✓	
Grammar		✓

The definitions of below-standard performance are attached to this report. Your child was (above)/below minimal standard on 4 of the 5 components of writing that we assessed.

VIII. INSTRUCTIONAL APPLICATIONS

Evaluation data, individual scores, and score distributions from district-wide assessment programs are generally beneficial to school administrators and curricular coordinators. The classroom teacher, however, seldom sees much useful application of the data for his or her own purposes -- designing writing activities for students. The activities described in this chapter suggest ways in which the teacher can make use of students' writing samples in the classroom, after the writing samples have been scored holistically and returned to the school. Three activities are presented. The first is appropriate for students in grades 4 through 12. The second is planned for younger writers (grades 3 through 5), although it could be adapted for older youngsters. The final activity is designed for students in grades 7 through 12. We hope that, as teachers become accustomed to working with holistic score results, these activities will spark ideas for additional teaching strategies.

ACTIVITY I (Grades 4 through 12)

1. Each student's paper will have a final score within the range 8 (high) to 2. Sort your papers into four stacks -- those scored 8, 6, 4, and 2. Put papers scored 3, 5, and 7 aside.

2. Skim through the separate stacks of papers and select one paper from each that you think is representative of the stack. You will have a "representative" (not the best) 8 paper, a representative 6 paper, and so on.

3. Duplicate a set of the four papers (score 8, 6, 4 and 2) for each member of your class. Be sure that a score is on each paper, but that no paper contains information that would identify the writer if he or she chooses to remain anonymous. If you think the alternate-numbered scores will confuse students, change the scores to 4, 3, 2, and 1.

4. Pass out sets of duplicated sample papers to students.

5. In a class discussion, have students compare the 8 paper with the 6 paper, the 6 paper with the 4 paper, and the 4 paper with the

2 paper. Treat each paired comparison separately and sequentially. For each comparison, ask students to identify the writing problems that the lower-scored paper has. List the problems that students identify on the blackboard.

6. Give students their own papers. Ask students to identify two or three of the listed problems they see in their own papers.

7. Assign a writing topic and have each student concentrate on improving the specific points he or she identified. The younger the student, the fewer should be the number of problems addressed in a single writing assignment. Subsequent assignments can focus on other writing problems the student has identified. More than one draft should be encouraged before the paper is seen by anyone else.

8. Critique each student's final draft, focusing only on those problems identified for special effort. Rather than correcting papers yourself, you might organize students into pairs, or small groups of no more than four, to read and discuss one another's papers.

This instructional activity has several advantages:

- Students will be motivated because they are focusing on problems they themselves see in their own writing.
- Students will aim for levels they need to attain and can achieve.
- Students will form the habit of looking critically at their own writing while they write.

ACTIVITY II (Grades 3 through 5)

1. Mix the papers that are returned to you and skim through them. Read each paper quickly but be sure to read the entire paper. Sort your papers into four stacks -- those scored 8, 6, 4, and 2. Put papers scored 3, 5, and 7 aside. Read through the 8 papers. Jot down one or two things most students receiving this score need to work on. Continue, analyzing 6s, 4s, and 2s, noting major points for each group of papers.

2. Select one major point of emphasis for each of the four score levels (8, 6, 4, and 2). You might decide, for example, that the 4s were characterized by weak organization; 6s would benefit from more attention to detail, and so on.

3. Re-read each paper carefully: read 3s, 5s, and 7s as well as 2s, 4s, 6s, and 8s. As you read each paper, classify it according to one of the four major writing problems you have identified. If a youngster has more than one of the four problems you selected, classify that student according to the simplest or lowest level problem. If, for example, a student needs to work on expressing complete thoughts, as well as developing paragraphs, he or she should concentrate on expressing ideas before thinking about organizing them.

4. Design a specific writing activity for each major writing problem you identified. For example, ask students whose writing evidences lack of detail to describe what they see, smell, and hear while sitting somewhere for five minutes: choose an environment that is rich in objects, colors, or other sensory stimuli -- such as the kitchen. Or assign the following topic to children who have difficulty with organization: "Write a story that tells about your last birthday. Begin with what happened in the morning and describe the important things that happened that day." You might suggest that youngsters make a "private" list of events that they can refer back to as they write their stories. This will help them organize their thoughts.

You will think of many more writing assignments that address specific problems. The major advantage of this activity is that it encourages differentiation. Why do we have groups of youngsters concentrate on different things in reading and mathematics and yet generally give an entire class the same writing assignment?

ACTIVITY III (Grades 7 through 12)

1. When you receive your students' scored essays, spend a few minutes reading all the papers, noting the score given to each. Then group the papers by major even-score level putting the odd-scored papers aside.

2. Make a list of about five or six positive writing characteristics of the four groups of papers. These characteristics should be generalizations -- for example, generally fluent, good sentence variety, etc. But the generalizations can be supported by illustrations from the actual papers. For each generalized characteristic, you should be able to refer to a specific paragraph or sentence to illustrate the characteristic.

3. Select one sample paper from each of the major score groups (8, 6, 4, and 2). This sample should not be the best of its respective group, but a middle-8, a middle-6, and so on. These sample papers should illustrate the generalizations you have developed for each of the score points.

4. Duplicate a set of sample papers for each student in your class. The papers should not be in order from high to low but rather mixed, and students should not know who wrote the papers or the scores they received.

5. Pass out the sets of papers to students. Tell students to read all the papers and then rank them from high to low based on their impressions of how well they are written.

6. In a directed discussion, elicit from the class five or six characteristics of each paper. Write these on the blackboard. You should draw out as many positive characteristics as you can. For example, one might be: the first paper generally shows evidence of appropriate organization; that is, it has a clear beginning or introduction, a middle or development phase, and a conclusion that rounds out the thesis or argument. Encourage discussion of the characteristics that students note -- both positive and negative ones.

7. Save these lists of characteristics and use them as scoring guides for future composition assignments.

This activity will encourage students to examine writing critically and will serve as an opportunity for exploring and communicating the standards you apply in judging writing.

APPENDIX A

REQUIREMENTS FOR WRITING TESTS (SECONDARY LEVEL)

A school district may use any instrument (state, commercial, or locally-developed) which assesses student writing by means of a writing sample as long as assurances are provided that the district will:

- Provide all students doing the writing sample an opportunity to use a dictionary.
- Use scoring procedures that assess student performance on the secondary level writing objectives set forth in Section 40.04 of the Regulations.
- Use scoring procedures that insure reliability of results.

Any commercial or locally-developed instrument used to assess writing skills must have the following characteristics:

- Each item is clear and concise.
- The directions are specific.
- The topic allows for a range of responses.
- The vocabulary is not too difficult to be understood by an average student.
- Each item indicates the way the student should proceed
- Each item is within the range of experience of all students.
- Each item is free of offensive sexual, cultural, racial, and ethnic content and stereotyping.

NOTE: When a district completes the Basic Skills Improvement Plan, the appropriate assurances and information for writing instrument(s) will be included. In completing and signing the Basic Skills Improvement Plan, a district will be assuring the Department of Education that these requirements will be met. No further approval process is required; there is no list of approved commercial instruments to assess writing skills.

APPENDIX B

SCORING WRITING COMPONENTS HOLISTICALLY

It is possible to vary holistic scoring methodology in order to evaluate individual elements or components of writing -- for example, organization or style. The procedures are as follows:

1. Identify the component of interest. These should be dimensions that will evidence a range of ability or performance.

2. Select training papers to illustrate the four levels of writing (represented by the score scale 4 to 1) for each component to be assessed.

3. Follow the same training and scoring procedures you us in conducting an holistic scoring session to judge writing for total effect. Introduce a single component, train the readers, and score the papers.

4. Present the second component, train the readers on that component, and score the papers.

5. Continue having the readers consider one component at a time until you have covered all the dimensions of writing that you plan to evaluate holistically.

When the scoring is complete, each paper will have a double score (the sum of two readers' scores) for each component you have addressed.

The standards-setting issues are identical to those discussed in Chapter VI of this manual. Minimal standards can be set be a panel prior to the scoring session, or after all the papers are scored. Since there will be a separate standard for each component, a single composite minimal standard may be set. Procedures for setting the composite standard on writing components scored holistically are identical to the guidelines presented on pages 70 and 71 of this manual.

APPENDIX C

PLANNING THE LENGTH OF THE SCORING SESSION

AND

ESTIMATING THE NUMBER OF READERS NEEDED

1. How many unique topics or separate writing assignments did students complete? _____ topics

2. Complete the following chart.

NUMBER OF PAPERS AND READINGS

Topic	Number of Papers	
1	_____	
2	_____	
3	_____	
Total	_____	(a)
Number of readings (a x 2)	_____	(b)

3. Assuming a reader can score one paper in 45 seconds, 80 papers in one hour, how many hours of actual reading time are needed ($\frac{b}{80}$)?

_____

Example:

Total papers (a)	600
Number of readings (b)	1200
$\frac{1200}{80}$ =	15 hours

4. Given the time you have available, how much reading time and how many readers do you need?

Allowing for training on[1] each topic and breaks, how much actual reading time do you have available? _____ (c)

How many hours of reading time do you need? (from question 3) _____ (d)

How many readers are needed $(\frac{c}{d})$? _____ Readers

Example: 2-1/2 hours (morning session) are available; 15 hours are needed (previous example)

$\frac{15}{2.5}$ = 6 readers

(If 1 1/2 hours are available, $\frac{15}{1.5}$ or 10 readers will be needed to score 600 papers.)

[1]Assume that it will probably take 30 to 40 minutes to train readers for the first exercise, with less time (approximately 20 to 25 minutes) for subsequent exercises. Plan for a 10-minute break in the morning and in the afternoon.

APPENDIX D

FOR THE READERS

The Process

You are to be part of a holistic scoring process that ensures that samples of student writing are judged as objectively as possible. You were selected to participate because you are experienced in dealing with a range of student writing.

When you use the holistic scoring method, you read a student's writing quickly and judge it for its total impact. Even if you have never scored holistically, you will find that your experience with student writing enables you to make the quick and definitive overall judgments required for holistic scoring. If a student paper contains a great many errors in spelling, syntax, or organization, those weaknesses will, of course, constitute part of your impression of the writing. But what is more important in these exercises is the way in which the student has expressed his or her ideas and opinions--the overall quality of the writing.

Even the best papers will not be flawless. When you consider the constraints we impose upon the students -- the time limitations, the tenseness of the testing situation -- you will realize why you and the other readers must not expect perfection or judge the writing sample against a preconceived "ideal" response. Instead, you should judge each test paper against the other test papers written for the same test administration. The purpose is to identify which students need further writing instruction.

The Training Session

In order to make a fair comparison of the test papers, you must, of course, have a "feel" for the kinds of responses and range of writing the exercise elicited. In the training part of the scoring session, the Chief Reader will present you with student papers similar to the types of papers

you will be scoring. As you and the other readers score and discuss these training papers, you will consider questions pertinent to the particular exercise.

For example, the questions to be resolved for one exercise might be: How closely should the directions be followed? What is considered "off the topic?" How important are spelling, punctuation, and grammar? The answers that the group agrees upon will become your criteria for scoring that exercise. In a conventional holistic scoring session, scoring criteria are not prescribed by the Chief Reader; rather, they are set during the training part of the scoring session. As the nature of the exercise and the types of student responses vary from one test administration to the next, so will the scoring criteria vary.

The Four-Point Rating Scale

You will score the test papers on a four-point scale, with 4 the highest score. Obviously, there is no middle score. One of the reasons why the four-point scale works well is that it forces readers away from a middle, or uncommitted, score. You must first decide whether a test paper, when compared with the other test papers, is in the upper-half (4-3) category or in the lower-half (2-1) category. Then you must decide whether that paper is written well enough to rate the highest score (4) or so poorly that it rates the lowest score (1).

New readers are often reluctant to grant any paper a 4. You must remember that the papers are being judged against each other, not against a "perfect" paper that might exist in your mind. Unfortunately, no reader, experienced or inexperienced, seems to need reassurance about giving out 2s and 1s; try to keep in mind that some papers are worth more than a 2 or a 1.

Score the Training Papers.

The Chief Reader and the Assistant Chief Reader will select test papers that will give you an idea of the types of responses to expect. You will read photocopies of the training papers and assign them scores. You may have doubts about how to score the first training papers, but after you have read six or seven, you should be able to compare papers and make judgments readily.

As you score the training papers, the Chief Reader will record the score distribution on a chalkboard. If, after scoring several training papers, you find that your scoring is inconsistent with that of the rest of the readers, you should adjust to the standards of the majority. Unless all readers judge the papers in the same way, the scoring procedure will be unreliable and unfair to the students.

Discussing Problems and Unusual Papers

During the reading, you might receive a test paper that is difficult to score because it is off-the-topic, illegible, or upsetting. You should always take such a paper immediately to the Chief Reader for consultation. Unusually creative papers--those that are really poems, short stories, or plays--should also be taken immediately to the Chief Reader. As "special" papers, they must be scored in a special way. In the training session, readers should set a policy for handling these papers. (If a more imaginative paper relates at all to the task assigned, you would probably be wise not to downgrade the student for allowing his or her imagination free rein. Just be certain that you never lose sight of the scoring criteria.)

When students have made little or no attempt to do the exercise and submit blank papers, papers with only a few, ineffective words, or papers that merely repeat the assignment, the paper should receive a score of 0. These papers must be taken to the Chief Reader as well.

The Scoring Session

As you finish a folder of papers, an aide will collect it and give you another folder. You are to read quickly and make a judgment based on your first impression of that piece of writing. The Chief Reader will monitor the scoring by reading various test papers after they have been scored. If the Chief Reader finds that several of you are losing sight of the scoring criteria or that there are too many score discrepancies, he or she may interrupt the scoring so that you may all be brought back into agreement.

Because each paper is assessed by two different readers, it will have two scores. If both scores match, the readers were in perfect agreement. If the score combination is a split of two adjacent numbers (4/3, 3/2, or 2/1), this split score is acceptable. However, discrepant scores from readers (combinations of either 4/2, 4/1, or 3/1) are not acceptable and must be resolved. Test papers with discrepant scores will be read and scored again by two different readers. If these readings also result in discrepant scores, the Chief Reader and the Assistant Chief Reader will assign the paper its final scores.

APPENDIX E

FOR THE AIDES

As an aide, you will assist the Chief Reader by preparing materials for the exercises to be scored. You will participate in the scoring session by distributing papers, recording scores, and following the specific instructions of the Chief Reader.

The Chief Reader may ask you to fold over and staple the cover of the test booklets so that the readers will not see the students' names. The Chief Reader may also ask you to separate test booklets so that several exercises can be scored at the same time by different groups.

Several Weeks Before the Scoring Session

1. Review all your responsibilities and meet with the Chief Reader to discuss them.

2. Make photocopies of each training paper given to you by the Chief Reader. The Chief Reader, the Assistant Chief Reader, and each of the readers will need their own copies of each training paper.

3. Make photocopies of the exercise (test question and directions) to be scored. (Make the same number of photocopies as you did in Step 2.)

4. For each exercise, mix students' test papers to that no reader has too many test papers from any one class or school. Put the original of each paper in folders in groups of 25, 20, or 15, in order to control their distribution more easily. Letter the papers in each folder beginning with the letter A.

5. Make scoring sheets and add to folders. You will need two scoring sheets for each folder of papers. These two forms should be identical, except that one should specify "First Reader's Signature" and the other,

"Second Reader's Signature." Be sure that the letters on each form correspond to the letters assigned to papers in each folder. If you have 20 papers in each folder, the correct 20 letters must be listed on each scoring sheet.

Folder #7	
Paper Letter	Score
A	
B	
C	
D	
E	
F	
G	
H	
I	
J	
K	
L	
etc.	
First Reader's Signature:	

Folder #7	
Paper Letter	Score
A	
B	
C	
D	
E	
F	
G	
H	
I	
J	
K	
L	
etc.	
Second Reader's Signature:	

The Day Before the Scoring Session

1. See that the room is set up according to the Chief Reader's instructions.

2. Place at each reader's seat and at the head table:
   - a supply of lined paper
   - two sharpened pencils

3. Be sure that the aides' tables have:
   - photocopies of the exercise to be scored
   - photocopies of the training papers
   - All student papers for the exercise that is about to be scored

4. Arrange students' test papers for distribution according to directions from the Chief Reader.

During the Scoring Session

When Readers are Trained

1. Distribute photocopies of the exercise (test questions and directions) when the Chief Reader tells you.

2. Distribute training papers when the Chief Reader tells you.

When Scoring Begins

1. According to instructions from the Chief Reader, distribute the papers to be scored.

2. Always place a folder of test papers on a reader's left and remove it from a reader's right.

3. Distribute and collect folders until all papers have been read twice. Do not allow a reader to sit idly, waiting for another folder of papers.

4. After papers in a folder have been read twice, the recording aide should check the scores. If the two scores are discrepant (such as 4/2, 4/1, or 3/1), the recording aide should remove the paper from the batch that has been scored satisfactorily and re-enter it into the distribution scheme, so that it can have a third and fourth reading by different readers. If the two new scores are also discrepant, the test paper should go to the Chief Reader for resolution. The recording aide should keep the Chief Reader informed of the number of discrepant scores.

5. If a reader hands you a test paper that has a score of zero, do not give that paper to a second reader but take it immediately to the Chief Reader.

6. When most of the papers have been read, give readers only two or three papers at a time to ensure that all readers finish at about the same time.

CHIEF READER AND AIDES' QUALITY CONTROL CHECKLIST

Room Arrangements

_____ Head table
_____ Readers' tables
_____ Aides' table(s)
_____ Chalkboard

Training Papers

_____ Photocopies in sufficient number
_____ Stacked in the order you will present them
_____ Master list of scores assigned to training papers

Student Papers

_____ Training papers returned to batch
_____ Mixed rather than grouped by class or school
_____ Organized for quick distribution
_____ Scoring sheets included
_____ Papers numbered

Materials for Each Reader

_____ A copy of each writing exercise
_____ A supply of lined paper
_____ Two sharpened pencils

_____ Arrangements for Coffee

APPENDIX F

SAMPLE RECORDING FORM

Analytic Scoring

STUDENT CODE _____

READER NUMBER __ __
1 2

	5	4	3	2	1	*
Organization	☐	☐	☐	☐	☐	☐
Clarity and Consistency of Purpose	☐	☐	☐	☐	☐	☐
Word Choice	☐	☐	☐	☐	☐	☐
Sentence Structure	☐	☐	☐	☐	☐	☐
Grammar	☐	☐	☐	☐	☐	☐

* Essay could not be scored in this category because

_____ writing is illegible

_____ essay is too brief to judge

_____ writer uses form other than essay (e.g., list)

_____ (other) _____

_____

BASIC SKILLS IMPROVEMENT POLICY

1980-81 ANNUAL REPORT ON BASIC SKILLS IMPROVEMENT PROGRAMS

System _____

System Code _____

Section 1 - Student Achievement of Minimum Standards

C. Secondary (7-12)[1]

(1) Skill (circle one): Reading Writing Mathematics

(2) Grade Level Reported (circle one): 7 8 9

(3) Evaluation Instrument(s) _____

(4) Description of Minimum Standard _____

(5) Students Evaluated for Achievement of Minimum Standards	All Students		Sex				Race/Ethnicity							
			F		M		B		H		W		O	
	#	%	#	%	#	%	#	%	#	%	#	%	#	%
(a) Students Achieving Minimum Standards														
(b) Students Not Achieving Minimum Standards														
(c) Total Evaluated		100%		100%		100%		100%		100%		100%		100%

(6) Students Exempted/Not Evaluated	All Students	Sex		Race/Ethnicity			
		F	M	B	H	W	O
	#	#	#	#	#	#	#
(a) Special Education Exemption							
(b) Limited English Ability Exemption							
(c) Other							
(d) Total Exempted/Not Evaluated							

[1] Comparable forms must be completed for early and later elementary standards.